

Bayeslösungen des Ausreißerproblems

von

FRIEDRICH GEBHARDT

1961

Bayes l ö s u n g e n
d e s
A u s r e i ß e r p r o b l e m s

Inaugural - Dissertation
zur Erlangung der Doktorwürde
der Naturwissenschaftlichen Fakultät
der Ludwig - Maximilians - Universität
München

vorgelegt von
Friedrich Gebhardt
aus Ober-Bellmannsdorf
Kreis Lauban

1. Juni 1961



1. Berichterstatter: Prof. Dr. H. Richter
2. Berichterstatter: Prof. Dr. K. Stein

Tag der mündlichen Prüfung: 26. Juli 1961

Inhaltsverzeichnis

	Seite
§ 1. Einleitung	4
§ 2. Übersicht über die Literatur zum Ausreißerproblem	8
§ 3. Statistische Entscheidungstheorie	19
§ 4. Konstanter Schaden	37
§ 5. Quadratischer Schaden und endlich viele Entscheidungen	44
§ 6. Quadratischer Schaden und unendlich viele Entscheidungen	50
Literaturverzeichnis	61
Lebenslauf	67



Digitized by the Internet Archive
in 2026

§ 1. Einleitung

Hat man unter den Beobachtungen $x_1, \dots, x_n$ der unabhängigen Zufallsgrößen $\xi_1, \dots, \xi_n$ eine oder mehrere, die von den übrigen stark abweichen, so liegt die Vermutung nahe, daß die zugehörigen Zufallsgrößen einer anderen Verteilung als die übrigen genügen, daß sie also sogenannte "Ausreißer" sind. Es sind verschiedentlich Tests vorgeschlagen worden, zu prüfen, welche der Zufallsgrößen als Ausreißer angesehen werden sollen. Wie in § 2 im einzelnen gezeigt wird, wurden zunächst zu verschiedenen, rein intuitiv gewonnenen Kriterien Sicherheitsschranken erster Art angegeben; welche Kriterien man benutzte, hing davon ab, ob man diese Schranken ausrechnen konnte. Erst ab 1950 wurden Gegenhypothese in die Betrachtung einbezogen und die Fehler zweiter Art berechnet. Für manche Tests wurde unter gewissen Einschränkungen die Optimalitätseigenschaft hergeleitet, bei vorgegebenem Fehler erster Art den Fehler zweiter Art zum Minimum zu machen.

In dieser Arbeit wollen wir mit den Mitteln der Testtheorie ein optimales Kriterium gewinnen; zu verschiedenen, intuitiv naheliegenden Schadensfunktionen bestimmen wir dabei die Bayeslösungen. Die statistische Entscheidungstheorie wird in § 3 bereitgestellt und dabei so verallgemeinert, wie sie in den folgenden Paragraphen benötigt wird. In § 4 werden zunächst nur Kriterien für die Feststellung gesucht, daß eine der Zufallsgrößen als Ausreißer anzusehen ist; der Schaden hängt dabei, wie es bei der Hypothesentestung im allgemeinen üblich ist, nur von der Hypothese H_j und der Entscheidung D_j ab, die in diesem Falle die Gestalt hat " H_j ist richtig", bei fester Hypothese und Entscheidung aber nicht mehr vom Ergebnis der Beobachtung. In unserem Zusammenhang bedeuten die Hypothesen H_j für $1 \leq j \leq n$, daß

g_j ein Ausreißer ist, H_0 , daß kein Ausreißer vorhanden ist. Die aus der Literatur bekannten Verfahren mit der oben genannten Optimalitätseigenschaft ergeben sich als Spezialfälle.

Im allgemeinen möchte man aber nicht nur wissen, ob ein Ausreißer vorhanden ist, sondern man möchte vor allem den Wert von Verteilungsparametern schätzen, etwa den Erwartungswert oder die Varianz. Zu diesem Zweck wird, wenn m den unbekannten Parameter und $h_t(g_1, \dots, g_n)$, $t \in T$, die zugelassenen Schätzfunktionen für diesen bedeuten, der quadratische Schaden $(h_t - m)^2$ eingeführt. Es handelt sich somit um eine Kombination von Hypothesentestung und Parameterschätzung. In § 5 wird der Fall untersucht, daß bei einer Gaußverteilung bekannter Varianz Eins der Erwartungswert gesucht wird. Ein etwa vorhandener Ausreißer soll dabei den gleichen Erwartungswert m und die ebenfalls bekannte Varianz σ^2 besitzen, und als Statistiken für m werden zugelassen:

$$h_0 = \frac{1}{n} \sum_{v=1}^n g_v \quad \text{und}$$

$$h_j = \frac{1}{n-1+\sigma^{-2}} \left(\frac{g_j}{\sigma^2} + \sum_{v \neq j} g_v \right) \quad (j=1, \dots, n);$$

h_j ($j \neq 0$) ist bekanntlich die effizienteste Statistik für den Fall, daß genau g_j ein Ausreißer ist mit der Varianz σ^2 .

In § 6 werden unter denselben Hypothesen alle meßbaren Funktionen $h(g_1, \dots, g_n)$ als Statistiken zugelassen. Aufgrund der Ergebnisse von § 3 können wir uns jedoch auf die Funktionen der Gestalt $h^*(x) = \bar{x} - a(x_1 - \bar{x}, \dots, x_n - \bar{x})$ beschränken, wobei wir zur Abkürzung x statt $x_1, \dots, x_n$ und $\bar{x}$ für das arithmetische Mittel der x_v schreiben. Ist dann $f_1(x, m)$ die gemeinsame Wahrscheinlichkeitsdichte von $g_1, \dots, g_n$ unter der Hypothese H_1 , so suchen wir die verallgemeinerten Bayeslösungen zur Vorbewertung $\varphi_0 = 1 - n\varphi$, $\varphi_1 = \dots = \varphi_n = \varphi$ der zusammenge-

setzten Hypothesen $H_0, H_1, \dots, H_n$, also eine Funktion $h^*(x) = \bar{x} - a(x_1 - \bar{x}, \dots, x_n - \bar{x})$, die unter allen Funktionen dieser Gestalt die von m unabhängige Risikosumme

$$\sum_i \varphi_i \int [h^*(x) - m]^2 f_i(x, m) dx_1 \dots dx_n$$

zum Minimum macht. Die Bayeslösung genügt in der Menge aller meßbaren Funktionen $h(x_1, \dots, x_n)$ der allgemeineren, in Satz (3.15) genannten Minimumsbedingung.

Für einige Werte von n, σ, φ wird das Risiko der Bayeslösung numerisch berechnet und mit dem Risiko eines der üblichen klassischen Verfahren verglichen. Es zeigt sich, daß beide Risiken von der gleichen Größenordnung sind, daß also die üblichen Verfahren im allgemeinen ausreichen.

Es soll hier noch ausdrücklich darauf hingewiesen werden, daß Ausreißertests natürlich empfindlich sind gegenüber einer Abweichung der wahren Verteilung der Zufallsvariablen von der der Rechnung zugrunde gelegten (meistens ist das die Normalverteilung) in den beiden Schwänzen, ferner darauf, daß die üblichen Verfahren im allgemeinen keine erwartungstreue Schätzung der unbekannten Parameter liefern.

Bezeichnungen.

Der Kürze und Übersichtlichkeit halber schreiben wir häufig $f(x)$ statt $f(x_1, \dots, x_n)$, ebenso dx statt $dx_1 \dots dx_n$. Gelegentlich werden die Summations- und Integrationsgrenzen weggelassen, wenn sich die Summation oder Integration über alle vorkommenden Werte der Indizes bzw. den ganzen Definitionsbereich der Integrationsvariablen erstrecken soll. Die Komponenten der Vektoren $\eta(x)$, $\kappa^*(x)$, $\delta^g(x)$ usw. bezeichnen wir mit $q_i(x)$, $r_i^*(x)$, $s_i^g(x)$ usw. Das Komplement einer Menge kennzeichnen wir

durch einen Querstrich. - Zahlen in eckigen Klammern verweisen wie üblich auf das Literaturverzeichnis.

Herrn Professor Dr. H.Richter, unter dessen Anleitung die vorliegende Arbeit entstand, gebührt besonderer Dank für zahlreiche Anregungen und kritische Durchsicht der Arbeit. Ebenso danke ich den Herren Dr. D.Bierlein und Dr. H.G.Kellerer für mehrere Verbesserungsvorschläge. Weiterhin gilt mein Dank Herrn Direktor Dr. H.Müller-Lutz und der Allianz Versicherungs-AG, München, für die Genehmigung, für numerische Rechnungen die elektronische Rechenanlage zu benutzen.

§ 2. Übersicht über die Literatur zum Ausreißerproblem

Seit einigen Jahrzehnten erscheinen immer wieder Arbeiten, die sich, von unterschiedlichen Voraussetzungen ausgehend und mit verschiedenen Methoden arbeitend, mit dem Ausreißerproblem beschäftigen. Die Arbeiten etwa aus der Zeit von 1925 bis 1950 lassen sich unter dem folgenden Schema zusammenfassen:

Als Kriterium dafür, ob die unabhängigen Beobachtungen $x_1, \dots, x_n$ übereinstimmend normal verteilt sind oder ob sich darunter eine oder mehrere anders verteilte befinden, schlagen die Verfasser der verschiedenen Arbeiten jeweils eine Statistik t vor mit der Anweisung: Ist die Realisation von t größer als die vorgegebene Konstante t_α , so wird die Hypothese angenommen, daß ein Ausreißer vorhanden ist. Der "kritische Wert" t_α wird so gewählt, daß unter der Nullhypothese, die die übereinstimmende Verteilung aller Beobachtungen annimmt, mit der Wahrscheinlichkeit $1 - \alpha$ die richtige Entscheidung getroffen wird. Über die Verteilung der Ausreißer werden keine Annahmen gemacht, daher ist die Bestimmung des Fehlers zweiter Art nicht möglich. Meistens werden die Werte t_α für verschiedene α berechnet, manchmal auch die Verteilungsfunktionen der vorgeschlagenen Statistiken t .

So benutzt 1925 J.O.Irwin [14] unter der Annahme, daß die Varianz σ^2 der gaußisch verteilten Beobachtungen bekannt sei, die Statistik $t = (u_n - u_{n-1})/\sigma$, wobei $u_1 \leq \dots \leq u_n$ die Rangstatistiken sind. Er tabelliert die Verteilungsfunktionen von $(u_n - u_{n-1})/\sigma$ und $(u_{n-1} - u_{n-2})/\sigma$ für verschiedene $n \leq 1000$.

L.H.C.Tippett [40] schlägt $t = (u_n - u_1)/\sigma$ vor und berechnet für mehrere $n \leq 1000$ die ersten vier Momente; E.S.Pearson und H.O.Hartley [12, 26, 27, 29] geben die

Verteilung dieser Statistik an.

Der Fehler erster Art der an sich naheliegenden Statistik $t = (u_n - \bar{u})/\sigma$ wird erst 1935 von A.T.McKay [19] untersucht; die kritischen Werte sind hier schwieriger zu erhalten. Er gibt eine Methode zu ihrer Berechnung an, K.R.Nair [22] und F.E.Grubbs [10] vereinfachen diese und berechnen mehrere Tafeln. Insbesondere ist bei F.E.Grubbs die Verteilungsfunktion von $(u_n - \bar{u})/\sigma$ für $2 \leq n \leq 25$ ausführlich tabelliert.

Darüber hinaus beschäftigt sich K.R.Nair wie vorher schon D.Newman [23] sowie E.S.Pearson und H.O.Hartley [30] und 1959 J.Pachares [24] und K.C.S.Pillai [34] mit dem Fall, daß die Varianz unbekannt ist und durch die Stichprobenstreuung s'^2 einer zweiten, unabhängigen Beobachtungsreihe ersetzt wird; alle Arbeiten enthalten Tafeln über die kritischen Werte und die Verteilungsfunktionen von $(u_n - u_1)/s'$ bzw. $(u_n - \bar{u})/s'$.

H.J.Godwin [9], E.S.Pearson und H.O.Hartley [13, 31] berechnen 1945 die Verteilungsfunktion von $t = \frac{1}{n} \sum |x_v - \bar{x}|$.

M.Halperin u.a. [11] geben 1955 kritische Werte an für die weitere Statistik $t = \max |x_v - \bar{x}|/s'$, wobei wiederum s'^2 die empirische Streuung einer zweiten, unabhängigen Beobachtungsreihe ist.

C.I.Bliss u.a. [1] berechnen bei n Stichproben zu je k Beobachtungen der Varianz 1 die 5%-Schranke für

$$t = \frac{\max_v (u_{vk} - u_{v1})}{\sum_{\mu=1}^n (u_{\mu k} - u_{\mu 1})} \quad \text{mit den innerhalb jeder Stichprobe}$$

geordneten Beobachtungen $u_{v1} \leq \dots \leq u_{vk}$.

Um mehrere Ausreißer zu entdecken, wird bei den bisher genannten Arbeiten empfohlen, den Test auf die restlichen Beobachtungen erneut anzuwenden, wenn eine Beobachtung als Ausreißer ausgeschieden wurde. Zwar hat man für die Wiederholungen keine Fehlerwahrscheinlichkeiten mehr,

doch ändert sich wenigstens der Fehler erster Art des gesamten Verfahrens dadurch nicht.

W.R.Thompson [39] führt 1935 die studentisierten Statistiken

$$t_i = (x_i - \bar{x}) / \sqrt{\frac{1}{n} \sum_{v=1}^n (x_v - \bar{x})^2} \quad (i = 1, \dots, n)$$

ein und verwirft diejenigen Beobachtungen x_i , für die $|t_i| > t_{\alpha/n}$ gilt. Die Konstante $t_{\alpha/n}$ wird bestimmt durch die Wahrscheinlichkeit unter der Nullhypothese

$p_0(|x_i - \bar{x}| > t_{\alpha/n}) = \alpha/n$. Der Fehler erster Art ist dann

kleiner oder gleich α . Wie E.S.Pearson und C.ChandraSekar [28], die im übrigen die Tafeln von W.R.Thompson erweitern, eingehend zeigen, kann diese Statistik für kleine n nicht zwei etwa gleichzeitig auftretende Ausreißer aufdecken, und wenn beide dem Betrage nach etwa gleich sind, so wird sogar keiner von beiden angezeigt. Zwar schlägt W.R.Thompson auch eine Statistik für mehrere Ausreißer $x_{v_1}, \dots, x_{v_k}$ vor, nämlich

$$t = \frac{\bar{x}_2 - \bar{x}}{\sqrt{\frac{1}{n} \sum_v (x_v - \bar{x})^2}} \quad \text{mit} \quad \bar{x}_2 := \frac{1}{k} \sum_i x_{v_i},$$

doch setzt diese Statistik voraus, daß man die Anzahl k der etwa vorhandenen Ausreißer im voraus kennt; denn testet man mit Hilfe dieses Verfahrens nacheinander auf 1, 2, 3, ... Ausreißer, so kennt man den Sicherheitskoeffizienten nicht mehr.

N.V.Smirnoff [37] benutzt 1941 die Statistik $t = (u_n - \bar{u})/s$ und berechnet verschiedene kritische Werte.

F.E.Grubbs [10] behandelt 1950 in der bereits oben zitierten Arbeit ebenfalls $t = (u_n - \bar{u})/s$ sowie die andere, damit gleichwertige Statistik

$$t = S_1^2 / S^2 \quad \text{mit} \quad S_1^2 := \sum_{v=2}^n (u_v - \bar{u}_1)^2, \quad \bar{u}_1 := \frac{1}{n-1} \sum_{v=2}^n u_v,$$

$S^2 := \sum_{v=1}^n (u_v - \bar{u})^2$. Außerdem untersucht er die entsprechend gebaute Statistik $S_{1,2}^2 / S^2$ als Test auf das Vorhandensein von zwei Ausreißern u_1 und u_2 und gibt kritische Werte an.

W.J.Dixon [4] ist der erste, der 1950 auch den Fehler zweiter Art untersucht, und zwar für fast alle bisher genannten Statistiken sowie für

$$t_{10} = \frac{u_2 - u_1}{u_n - u_1}, \quad t_{11} = \frac{u_2 - u_1}{u_{n-1} - u_1}, \quad t_{12} = \frac{u_2 - u_1}{u_{n-2} - u_1},$$

$$t_{20} = \frac{u_3 - u_1}{u_n - u_1}, \quad t_{21} = \frac{u_3 - u_1}{u_{n-1} - u_1}, \quad t_{22} = \frac{u_3 - u_1}{u_{n-2} - u_1}.$$

Dabei benutzt er die Nullhypothese

H_0 : "alle Variablen sind nach $N(m, \sigma^2)$ verteilt"

und die Gegenhypothesen

H_1 : "eine der Variablen ist nach $N(m + \lambda \sigma, \sigma^2)$ verteilt" oder

H_2 : "eine der Variablen ist nach $N(m, \lambda^2 \sigma^2)$ verteilt".

In manchen Fällen werden auch zwei Ausreißer betrachtet.

Die Fehler zweiter Art werden für einige Werte von λ experimentell bestimmt. Es stellt sich heraus, daß bei be-

kanntem σ die Statistik $(u_n - \bar{u})/\sigma$ zu festem Fehler erster Art den kleinsten Fehler zweiter Art besitzt, daß je-

doch $(u_n - u_1)/\sigma$ in dieser Hinsicht nur wenig schlechter ist. Bei unbekanntem σ ist t_{10} für kleine n gut, t_{22}

für $n > 15$ und für $8 \leq n \leq 13$ vermutlich t_{21} . Wenn

höchstens an einer Seite Ausreißer auftreten können, ist

S_1^2 / S^2 etwa eben so gut wie t_{10} ; für große n dürfte

$S_{1,2}^2 / S^2$ nützlich sein, doch waren damals die kritischen

Werte nur für $n \leq 11$ bekannt.

In einer weiteren Arbeit veröffentlicht W.J.Dixon [5] für die sechs Statistiken $t_{10}, t_{11}, t_{12}, t_{20}, t_{21}, t_{22}$ für $n \leq 30$ die kritischen Werte und in einigen Fällen die Wahrscheinlichkeitsdichten.

H.A.David [3] tabelliert 1956 die Wahrscheinlichkeit, einen Ausreißer mit einem nach rechts verschobenen Mittelwert zu finden zu einer vorgegebenen Sicherheitschranke von 5% unter der Nullhypothese.

In den Jahren ab 1952 erscheinen mehrere Arbeiten, die in folgendem Sinne eine optimale Statistik suchen:

Unter der Nullhypothese H_0 seien n Stichproben nach der gleichen Gaußverteilung verteilt; bei den Gegenhypothesen H_1 besteht die i -te Stichprobe aus ebenfalls normal verteilten Ausreißern mit abweichendem Mittelwert oder abweichender Varianz. Die Annahmen über die Größe der Stichproben und darüber, ob Erwartungswert und Varianz unter der Nullhypothese bekannt sind, sind in den einzelnen Arbeiten verschieden. Es dürfen die Entscheidungen D_1 , daß H_1 richtig ist, getroffen werden. Gesucht ist ein Entscheidungsverfahren mit der Eigenschaft:

(2.1) Der Fehler zweiter Art wird zum Minimum gemacht unter den Einschränkungen:

- a) Unter H_0 wird D_0 mit der vorgegebenen Wahrscheinlichkeit $1 - \alpha$ angenommen,
- b) das Verfahren ist symmetrisch in allen Stichprobenvariablen,
- c) es ist invariant gegenüber der Addition einer Konstanten zu allen Stichprobenvariablen,
- d) es ist invariant gegenüber der Multiplikation aller Stichprobenvariablen mit einer positiven Konstanten.

Als Lösung wird von den Verfassern jeweils eine Statistik der Gestalt $t = \max_j t_j$ angegeben mit der Vorschrift: Ist $t \leq t_\alpha$ mit der durch (2.1 a) bestimmten Konstanten t_α , so soll D_0 angenommen werden, andernfalls die durch $t_j = \max_v t_v$ bestimmte Entscheidung D_j .

So betrachtet E. Paulson [25] n Stichproben gleichen Umfanges k mit übereinstimmender, aber unbekannter Varianz. Unter der Gegenhypothese H_1 hat die i -te Stichprobe einen um d größeren Erwartungswert als die restlichen Stichproben. D_0 wird angenommen, wenn

$$t = (\bar{x}_M - \bar{x})/s \leq t_\alpha \quad \text{mit} \quad \bar{x}_v = \frac{1}{k} \sum_x x_{vx}, \quad \bar{x}_M = \max_v \bar{x}_v.$$

Ist $t > t_\alpha$, so wird D_M angenommen.

D.R. Truax [41] behandelt ebenfalls n Stichproben zu je k Beobachtungen, aber jede Stichprobe darf einen anderen, unbekannten Erwartungswert besitzen. Die Bedingung (2.1 c) wird dahingehend verschärft, daß das Verfahren auch bei Addition verschiedener Konstanten zu den verschiedenen Stichproben invariant sein soll. Die Gegenhypothese H_1 lautet: Die i -te Stichprobe besitzt eine größere Varianz als die übrigen. D_0 wird angenommen, wenn

$$t = \frac{s_M^2}{\sum_v s_v^2} \leq t_\alpha \quad \text{mit} \quad s_v^2 = \frac{1}{k-1} \sum_{x=1}^k (x_{vx} - \frac{1}{k} \sum_{\mu=1}^k x_{v\mu})^2, \\ s_M^2 = \max_v s_v^2.$$

Ist $t > t_\alpha$, so wird D_M angenommen. Die Werte t_α sind z.B. bei E.S. Pearson und H.O. Hartley [43] tabelliert.

K.V. Ramachandran und C.G. Khatrı [35] lassen bei n Stichproben mit je k Beobachtungen die Hypothesen über die Erwartungswerte m_v

$$H_0: m_1 = \dots = m_n := m,$$

$$H_{ij}: m_v = m \quad \text{für} \quad v \neq i, j, \quad m_i = m - d, \quad m_j = m + d \\ (i, j = 1, \dots, n; i \neq j)$$

zu und die Entscheidungen

D_0 : " H_0 ist richtig ",

D_{ij} : " H_{ij} ist richtig ".

D_0 wird angenommen, wenn

$$t = (\bar{x}_M - \bar{x}_L) / s \leq t_\alpha \quad \text{mit} \quad \bar{x}_v = \frac{1}{k} \sum_{x=1}^k x_{vx},$$

$$\bar{x}_M = \max_v \bar{x}_v, \quad \bar{x}_L = \min_v \bar{x}_v;$$

ist $t > t_\alpha$, so wird D_{LM} angenommen. Die Werte t_α lassen sich aus Tafeln für die Spannweite von Gaußvariablen berechnen, z.B. [30, 43].

A.Kudo [17] verallgemeinert die von K.R.Nair und F.E. Grubbs behandelten Statistiken und beweist die Optimalitätseigenschaft (2.1) für den folgenden Fall. Gegeben seien drei Stichproben:

$x_v^{(1)}$ ($v=1, 2, \dots, n_1$) sei nach $N(m_v, \sigma^2)$ verteilt,

$x_v^{(2)}$ ($v=1, 2, \dots, n_2$) sei nach $N(m', \sigma^2)$ verteilt,

$x_v^{(3)}$ ($v=1, 2, \dots, n_3$) sei nach $N(m'', \sigma^2)$ verteilt;

m_v, m', m'' und σ sind unbekannt. Die Hypothesen lauten:

H_0 : $m_1 = \dots = m_{n_1} = m'$,

H_1 : $m_v = m'$ für $v \neq i$, $m_i = m' + d > m'$ ($i=1, 2, \dots, n_1$).

D_M (H_M ist richtig) wird für $M \neq 0$ angenommen, wenn mit den Bezeichnungen

$$\bar{x}_{12} = \frac{1}{n_1 + n_2} (\sum x_v^{(1)} + \sum x_v^{(2)}), \quad \bar{x}_3 = \frac{1}{n_3} \sum x_v^{(3)}, \quad x_M = \max_v x_v$$

die Ungleichung gilt

$$t = \frac{x_M - \bar{x}}{\sqrt{\sum (x_v^{(1)} - \bar{x}_{12})^2 + \sum (x_v^{(2)} - \bar{x}_{12})^2 + \sum (x_v^{(3)} - \bar{x}_3)^2}} > t_\alpha;$$

ist $t \leq t_\alpha$, so wird D_0 angenommen.

Bei allen genannten Statistiken treten die Parameter des

Ausreißers nicht explizit auf, doch hängt natürlich der Fehler zweiter Art von ihnen ab.

R.Doornbos und H.J.Prins [8] verallgemeinern den von E.Paulson angegebenen Test auf Stichproben ungleichen Umfanges, doch besitzt er dann nicht mehr die Optimalitätseigenschaft (2.1). Außer für die Gaußverteilung stellen sie auch für einige andere Verteilungen Ausreißertests auf, nämlich in zwei vorhergehenden Arbeiten [6, 7] für die Gammaverteilung und in dieser für einige diskrete Verteilungen. Bei der Poissonverteilung besitzt der Test die Eigenschaft (2.1), wobei die Bedingungen (2.1 c,d) hier natürlich hinfällig sind.

J.Pfanzagl [32] formuliert das Problem allgemeiner. Zu einem σ -Körper über M seien die Wahrscheinlichkeitsmaße $\mu_0, \mu_1, \dots, \mu_n$ gegeben. Die Hypothese H_i ($i=0, 1, \dots, n$) nimmt μ_i als wahr an. Wird die Nullhypothese H_0 , für die der Sicherheitskoeffizient α vorgeschrieben ist, verworfen, so soll zwischen den übrigen Hypothesen, die alle "gleich wahrscheinlich" seien, entschieden werden. Gesucht wird ein Funktionensystem $\Phi := \{\phi_0(x), \dots, \phi_n(x)\}$ mit

$$\phi_\nu(x) \geq 0, \quad \sum_{\nu=0}^n \phi_\nu(x) = 1 \quad \text{für alle } x, \quad \int \phi_0(x) d\mu_0 = 1-\alpha,$$

so daß $\sum_{\nu=1}^n \int \phi_\nu(x) d\mu_\nu$ zum Minimum gemacht wird. Alle Integrale erstrecken sich über den ganzen Raum M . Es wird also nicht verlangt, daß unter jeder der Gegenhypothesen $H_1, \dots, H_n$ die Wahrscheinlichkeit einer richtigen Entscheidung maximal wird, sondern die Summe dieser Wahrscheinlichkeiten. Φ^* heißt "t & k-optimal" (test- und klassifikationsoptimal, weil erst H_0 gegen die übrigen H_i getestet und im Falle der Ablehnung zwischen diesen entschieden wird), wenn es kein Φ gibt mit

$$\int \phi_0(x) d\mu_0 \geq \int \phi_0^*(x) d\mu_0, \quad \sum_{v=1}^n \int \phi_v(x) d\mu_v > \sum_{v=1}^n \int \phi_v^*(x) d\mu_v.$$

Es gilt der

Satz: ϕ^* ist genau dann t & k-optimal, wenn für eine Konstante c μ -fast gilt

$$\phi_0^*(x) = 0, \quad \text{wenn} \quad c \cdot f_0(x) < \max[c \cdot f_0(x), f_1(x), \dots, f_n(x)],$$

$$\phi_v^*(x) = 0, \quad \text{wenn} \quad f_v(x) < \max[c \cdot f_0(x), f_1(x), \dots, f_n(x)]$$

$$(v = 1, 2, \dots, n)$$

mit den Wahrscheinlichkeitsdichten $f_v(x)$ der Maße μ_v bezüglich eines Maßes μ .

Die bei der Konstruktion von ϕ^* auftretende Funktion

$$t(x) := \frac{\max[f_1(x), \dots, f_n(x)]}{f_0(x)} \quad \text{heißt Entscheidungs-}$$

kriterium. Die Konstante c wird so gewählt, daß

$$t(x) \int_{t(x) < c} d\mu_0 \leq 1 - \alpha \quad \text{und} \quad t(x) \int_{t(x) \leq c} d\mu_0 \geq 1 - \alpha \quad \text{ist. Gibt es}$$

zu jedem α Tests, die gleichmäßig trennscharf gegen die zusammengesetzte Hypothese $H = \bigcup_{i \neq 0} H_i$ sind, so gibt es auch einen t & k-optimalen, gleichmäßig trennscharfen Test.

Angewandt auf das Ausreißerproblem, bei dem H_i ($i \neq 0$) die Beobachtung x_i zum Ausreißer erklärt mit der Dichte $\psi_1(x_i)$ anstelle von $\psi_0(x_i)$, erhält man

$$t(x) = \max_v \frac{\psi_1(x_v)}{\psi_0(x_v)}.$$

J. Pfanzagl rechnet $t(x)$ für verschiedene Fälle aus. Sind z.B. bei n Stichproben ungleichen Umfanges k_v unter der Nullhypothese alle Variablen nach $N(a_0, \sigma^2)$ verteilt, unter den Gegenhypothesen aber die Variablen der aus Ausreißern bestehenden Stichprobe nach $N(a_1, \sigma^2)$, so ist $t(x)$ äquivalent mit

$$t'(x) = \max_v \left[k_v \left(\bar{x}_v - \frac{a_0 + a_1}{2} \right) \right] \quad (a_1 > a_0),$$

§ 3. Statistische Entscheidungstheorie *)

Wir betrachten zunächst endlich viele Hypothesen und eine beliebige Menge von Terminalentscheidungen. Über einer Menge M mit den Elementen $x \in M$ sei ein σ -Körper $\mathcal{A}$ gegeben; über $\mathcal{A}$ seien endlich viele Wahrscheinlichkeitsmaße $\mu_1, \dots, \mu_n$ definiert mit den Dichten $f_i(x)$ ($i=1, \dots, n$) bezüglich eines Maßes μ . Zur Vereinfachung der Formulierungen nehmen wir ohne wesentliche Einschränkung an, daß das Maß μ vollständig ist.

Die Menge der Terminalentscheidungen sei T . Es wird zugelassen, daß unter der Hypothese H_i , die $f_i(x)$ als richtig annimmt ($i=1, \dots, n$), der eintretende Schaden d_i außer von der Entscheidung $t \in T$ auch noch von der Beobachtung $x \in M$ abhängt: $d_i = d_i(x; t)$. Wir fassen $d_1, \dots, d_n$ zum Schadensvektor $\mathcal{J}(x; t)$ zusammen. Allgemein bezeichnen wir die Komponenten der n -dimensionalen Vektoren $\eta(x)$, $\eta^*(x)$, $\mathcal{J}^s(x)$ usw. resp. mit $q_i(x)$, $r_i^*(x)$, $s_i^s(x)$ usw., $i=1, \dots, n$.

Letzten Endes interessiert man sich für Entscheidungsfunktionen $\tau(x)$, also Abbildungen von M in T , mit bestimmten optimalen Eigenschaften. Diese Eigenschaften betreffen jedoch nicht $\tau(x)$ selbst, sondern nur die zugehörigen Schadensfunktionen $\mathcal{J}(x; \tau(x))$. Wir werden deshalb

*) Die Ausführungen dieses Paragraphen gehen zum Teil auf eine Vorlesung zurück, die H. Richter im Sommersemester 1955 an der Universität München gehalten hat. An Literatur sei auf die im Teil b) des Literaturverzeichnisses genannten Arbeiten hingewiesen, insbesondere auf den Bericht von D.V. Lindley [48], auf das Buch von A. Wald [51] und auf die Arbeit über mehrstufige Tests von H. Richter [50].

von vornherein von einer Menge S von Schadensfunktionen ausgehen, die bestimmte, in den Definitionen (3.1), (3.2) bzw. (3.11) zu nennende Eigenschaften besitzen soll. Die Entscheidungsfunktionen spielen dabei nur insoweit eine Rolle, als es zu jeder Funktion $\phi(x) \in S$ eine Funktion $\tau(x)$ geben soll mit $\phi(x) = \mathcal{J}(x; \tau(x))$.

(3.1) Definition. S sei eine nicht leere Menge von n -dimensionalen, reellen Funktionen $\phi(x)$ mit den Eigenschaften:

a) $s_i(x)$ ist μ_i -integabel ($i=1,2,\dots,n$).

b) Ist $A \in \mathcal{A}$ und $\phi^1, \phi^2 \in S$, so ist

$$\phi := \begin{cases} \phi^1 & \text{auf } A \\ \phi^2 & \text{auf } \bar{A} \end{cases}$$

ebenfalls ein Element aus S .

c) Konvergieren für eine Folge $\phi^1, \phi^2, \dots$ die Produkte $s_i^g f_i$ ($i=1, \dots, n$) μ -fast überall, so gibt es eine Funktion $\phi \in S$ mit

$$s_i f_i = \lim_{g \rightarrow \infty} s_i^g f_i \quad \mu\text{-fast überall.}$$

d) $\phi(x)$ läßt sich darstellen in der Gestalt

$$\phi(x) = \mathcal{J}(x; \tau(x)) .$$

Dann heißt S "vom Typ A (über $\mathcal{A}$)"; $\tau(x)$ heißt "Entscheidungsfunktion" oder "Strategie".

Die in dieser Definition genannten Eigenschaften reichen zum Beweis der meisten folgenden Sätze aus; teilweise benötigen wir jedoch stärkere Voraussetzungen, nämlich, wie wir sogleich präzisieren werden, die Kompaktheit einer S erzeugenden Funktionenmenge Q und die Tatsache, daß alle durch Q erzeugten, meßbaren Funktionen zu S gehören. Wir geben nun zunächst die Definition und anschließend einige Erläuterungen.

(3.2) Definition. Q sei eine nicht leere Menge von n -dimensionalen, durch einen Parameter $u \in U$ gekennzeichneten Funktionen $\varphi(x;u)$ mit den Eigenschaften:

a) q_i ist μ_i -integrabel ($i=1, \dots, n$).

b) Q ist kompakt bei der durch den Abstand

$$\Delta(f, f^*) = \sup_x \left(\frac{1}{g(x)} \sum_1 |h_i(x) - h_i^*(x)| \right)$$

induzierten Topologie mit einer passenden, für alle i μ_i -integrablen Gewichtsfunktion $g(x) > 0$.

c) Zu $\varphi(x;u)$ gibt es eine Strategie $\tau_u(x)$ mit $\vartheta(x; \tau_u(x)) = \varphi(x;u)$.

Ferner sei S die Menge aller Funktionen $\delta(x)$, für die gilt:

d) s_i ist μ_i -meßbar ($i=1, \dots, n$).

e) $\delta(x)$ läßt sich darstellen in der Gestalt

$$\delta(x) = \varphi(x; v(x)) .$$

Dann heißt S " vom Typ B (über δ) ".

(3.2*) Anmerkung. (3.2 b) ist sicher dann erfüllt, wenn

a) Q nur endlich viele Elemente enthält oder

b) $\varphi(x;u)$ für $\varphi \in Q$ nicht von x abhängt und Q eine kompakte Teilmenge des R^n bildet.

Ist S vom Typ B, so lassen sich die Funktionen $\delta \in S$ wegen (3.2 e, c) ebenfalls in der Gestalt $\delta(x) = \vartheta(x; \tau(x))$ schreiben.

Die in (3.2) eingeführte Menge Q kann beispielsweise die Menge aller Funktionen $\vartheta(x;t)$ sein; dann kann man U mit T identifizieren. Dieser Fall wird in den §§ 4 und 5 eintreten. In § 6 werden wir jedoch vorübergehend nicht alle Entscheidungen aus T (die dort durch eine reelle Zahl a gekennzeichnet werden) zulassen, sondern nur diejenigen, für die $|a| \leq A(x)$ gilt mit einer passenden Funktion $A(x)$;

die Menge der zugelassenen Entscheidungen hängt also von $x \in M$ ab. Wir werden deshalb dort $\varphi(x; u) = \mathcal{J}(x; u \cdot A(x))$ setzen, $|u| \leq 1$.

Im übrigen kann es durchaus vorkommen, daß verschiedene Funktionen $\tau_1(x)$, $\tau_2(x)$ die gleiche Funktion $\mathcal{G} \in S$ erzeugen; es ist auch zulässig und stört die weiteren Entwicklungen nicht, wenn $\varphi(x; u_1) = \varphi(x; u_2)$ ist für $u_1 \neq u_2 \in U$.

Wir zeigen nun, daß (3.2) tatsächlich stärkere Forderungen stellt als (3.1)

(3.3) Satz. Ist S vom Typ B, dann ist S auch vom Typ A.

Beweis. Aufgrund der Konstruktion von S aus Q sind die Bedingungen (3.1 b, d) trivialerweise erfüllt. - Wegen der Kompaktheit von Q gibt es endlich viele Funktionen $\varphi^\lambda(x) := \varphi(x; u_\lambda) \in Q$, $\lambda = 1, 2, \dots, l$, so daß zu jedem φ aus Q ein λ existiert mit $\Delta(\varphi, \varphi^\lambda) \leq 1$. Für ein vorgegebenes $\mathcal{G} \in S$ und festes x^* gilt dann $\mathcal{G}(x^*) = \varphi(x^*; u^*)$ mit geeignetem $u^* \in U$ und daher

$|s_1(x^*)| \leq g(x^*) + |q_1^\lambda(x^*)|$ mit geeignetem, von x^* und $\mathcal{G}$ abhängigem $\lambda \leq l$. Somit gilt schließlich für alle x und sogar unabhängig von $\mathcal{G}$

$$|s_1(x)| \leq g(x) + \sum_{\lambda=1}^l |q_1^\lambda(x)|.$$

Die rechte Seite ist μ_1 -integrabel, also gilt (3.1 a). - Zum Nachweis von (3.1 c) mögen die Produkte $s_1^g f_i$ auf dem Komplement $\bar{N}$ einer μ -Nullmenge N konvergieren, der Limes ist also auf $\bar{N}$ μ -meßbar. Für festes x^* ist nach (3.2 e) $\mathcal{G}^g(x^*) = \varphi(x^*; u_g(x^*))$. Wegen (3.2 b) besitzt die Folge der Funktionen $\varphi(x; u_g(x^*))$ einen Häufungspunkt $\varphi(x; v(x^*)) \in Q$. Für die Funktion $\varphi(x; v(x))$ gilt $q_1(x; v(x)) f_i(x) = \lim_{g \rightarrow \infty} s_1^g(x) f_i(x)$ μ -fast überall, woraus auch die μ_1 -Meßbarkeit von $q_1(x; v(x))$ folgt; w.z.b.w.

(3.4) Definition. Die Komponenten s_i von δ seien μ_1 -integrabel. Dann heit $r(\delta) := (r_1(\delta), \dots, r_n(\delta))$ mit $r_i(\delta) := E_1(s_i) = \int_M s_i(x) f_i(x) d\mu$ " Risikovektor zu δ ".

(3.5) Hilfssatz. Ist μ ein atomloses Ma, sind $f_{ij}(x)$ μ -integrable Funktionen ($i=1, 2, \dots, n; j=1, 2$) und ist $r_{ij} := \int_M f_{ij}(x) d\mu$, so gibt es zu jedem θ mit $0 < \theta < 1$ eine μ -mebare Menge $A \subset M$ mit $\int_A f_{i1}(x) d\mu + \int_{A^c} f_{i2}(x) d\mu = \theta r_{i1} + (1 - \theta) r_{i2}$ gleichzeitig fr alle i .

Der Beweis findet sich z.B. bei D.Blackwell [45]. Aus dem Hilfssatz folgt sofort der folgende Satz.

(3.6) Satz. Ist S vom Typ A und μ atomlos, dann ist $\mathcal{R} := \{r(\delta) : \delta \in S\}$ konvex.

(3.7) Satz. Unter den Voraussetzungen (3.1 a, b, c) fr S gilt:
 $\mathcal{R} = \{r(\delta) : \delta \in S\}$ ist kompakt.

Der Beweis dieses Satzes wurde von D.Blackwell [45] unter Benutzung der hier angegebenen Voraussetzungen gefhrt, obgleich er in der Formulierung eine strkere Voraussetzung benutzt, nmlich (3.2* b) nebst (3.2 d, e) statt (3.1 c).

(3.8) Definition. S sei vom Typ A.

- a) $\delta^* \in S$ heit "Minimaxlsung", falls $\max_i r_i(\delta^*) \leq \max_i r_i(\delta)$ fr alle $\delta \in S$.
 b) Das Zahlensystem $\varphi := (\varphi_1, \dots, \varphi_n)$ heit

"Vorbewertung", sofern $\varphi_i \geq 0$ ist für alle i und $\sum_i \varphi_i = 1$.

- c) Die Schadensfunktion $\delta^* \in S$ wird "Bayeslösung zur Vorbewertung φ " genannt, wenn
- $$\sum_i \varphi_i r_i(\delta^*) \leq \sum_i \varphi_i r_i(\delta) \quad \text{für alle } \delta \in S.$$

Aus der Konvexität und Kompaktheit von $\mathcal{R}$ ergibt sich unmittelbar der folgende Satz.

- (3.9) Satz. S sei vom Typ A und μ atomlos. Dann existiert zu jeder Vorbewertung φ mindestens eine Bayeslösung $\delta^* \in S$. Ferner existiert mindestens eine Minimallösung δ^* , und alle Minimallösungen sind Bayeslösungen zu einer gemeinsamen Vorbewertung φ^* , wobei $\varphi_i^* = 0$ gilt für alle i mit $r_i^* < \max_j r_j^*$, $r_j^* := r_j(\delta^*)$.

Der folgende Satz dient dazu, die Gestalt der Bayeslösungen zu finden.

- (3.10) Satz. Ist S vom Typ B, so ist $\delta^*(x) \in S$ genau dann Bayeslösung zur Vorbewertung φ , wenn μ -fast überall gilt

$$(*) \quad \sum_i \varphi_i s_i^*(x) f_i(x) = \inf_{\delta \in S} \sum_i \varphi_i s_i(x) f_i(x).$$

Beweis. Der Beweis ist analog zu dem des entsprechenden Satzes von H. Richter [50]. Wegen der Kompaktheit von Q gibt es eine abzählbare Folge $q^1, q^2, \dots$ aus Q , die in Q überall dicht liegt. Aus (3.1 b) folgt sofort für jede Bayeslösung $\delta^*(x)$ und $s = 1, 2, \dots$

$$\sum_i \varphi_i s_i^*(x) f_i(x) \leq \sum_i \varphi_i q_i^s(x) f_i(x) \quad \text{für } x \in \bar{N}_s,$$

wobei N_s eine μ -Nullmenge ist. Daher gilt für alle $x \in \bar{N}$

$$\text{mit } N = \bigcup_s N_s \cup \{g(x) = \infty\}$$

$$\begin{aligned}
\sum_i \varphi_i s_i^*(x) f_i(x) &= \inf_{\varphi} \sum_i \varphi_i q_i^{\varphi}(x) f_i(x) \\
&= \inf_{u \in U} \sum_i \varphi_i q_i(x; u) f_i(x) \\
&= \inf_{\delta \in S} \sum_i \varphi_i s_i(x) f_i(x) \quad (x \in \bar{N}) .
\end{aligned}$$

Umgekehrt folgt aus (*) sofort durch Integration über M , daß δ^* eine Bayeslösung zur Vorbewertung φ ist, w.z.b.w.

Bemerkung. Für Satz (3.10) genügt nicht die Voraussetzung, daß S vom Typ A ist. Wie man an Beispielen zeigen kann, braucht das Infimum dann nicht meßbar zu sein, und wenn es meßbar ist, braucht es von keiner Funktion $\delta^* \in S$ μ -fast angenommen zu werden.

Wir werden in den folgenden Paragraphen n unabhängige Gaußvariable $g_1, \dots, g_n$ mit übereinstimmendem, aber unbekanntem Erwartungswert m bei verschiedenen Hypothesen über die Varianzen betrachten. In den neuen Variablen

$$a_0 := \bar{g} = \frac{1}{n} \sum_v g_v, \quad a_v := g_v - \bar{g} \quad (v = 1, \dots, n-1)$$

sind die Wahrscheinlichkeitsdichten von $a_0, \dots, a_{n-1}$ Funktionen der Gestalt $f_i = f_i(y_0 - m, z)$, wenn wir zur Abkürzung z statt $y_1, \dots, y_{n-1}$ schreiben. Letzten Endes wollen wir uns auf Entscheidungsfunktionen der Gestalt $\tau(z)$ beschränken, die also entsprechend der Bedingung (2.1 c) nicht von y_0 abhängen. Da auch die Schadensvektoren die Gestalt $\delta(y_0 - m, z; t)$ haben werden, sind dann die Risiken

$$\int d_1(y_0 - m, z; \tau(z)) f_i(y_0 - m, z) dy_0 dz$$

unabhängig von m . Die genannte Beschränkung auf Entscheidungsfunktionen $\tau(z)$ wird dadurch gerechtfertigt werden, daß es Funktionen der Gestalt $\delta(y_0 - m, z; \tau(z))$ gibt mit einer in Satz (3.15) genannten Optimaleigenschaft, die

besagt, daß diese Funktionen in Verallgemeinerung von Definition (3.8) als Bayeslösungen bezüglich zusätzlicher gleicher Vorbewertung aller Parameterwerte m angesehen werden können.

Wir wollen jetzt die eben skizzierten Gedanken ausführen. Dazu sei im folgenden $M = (R^1, M')$, $x = (y, z)$ mit $y \in R^1$ und $z \in M'$, $\mathcal{A}'$ ein σ -Körper mit vollständigem, atomlosem Maß μ' über M' und L das Lebesguesche Maß auf R^1 ; $\mathcal{A}$ sei der zur Vervollständigung μ des Maßes $L \times \mu'$ gehörige σ -Körper über M .

Wir betrachten die zusammengesetzten Hypothesen H_i ($i=1, \dots, n$); zu H_i gehöre die Wahrscheinlichkeitsdichte $f_i(y - m, z)$ mit dem Parameter $m \in R^1$. Für festes $t \in T$ hänge der Schaden von y , z und m nur in der Gestalt $\vartheta(y - m, z; t)$ ab.

Da an die Stelle der Einzelhypothese H_i jetzt eine von $m \in R^1$ abhängige, zusammengesetzte Hypothese tritt, benötigen wir eine Modifikation der früheren Definitionen (3.2) und (3.8).

(3.11) Definition. Q sei eine nicht leere Menge von n -dimensionalen, reellen, durch einen Parameter $u \in U$ gekennzeichneten Funktionen $q(y, z; u)$ mit den Eigenschaften:

- a) $q_i(y, z; u) f_i(y, z)$ ist μ -integrabel ($i=1, \dots, n$).
- b) Q ist kompakt bei der durch den Abstand

$$\Delta^*(f, f^*) = \sup_{y, z} \left(\frac{1}{g(z)} \sum_i |h_i(y, z) - h_i^*(y, z)| \right)$$

induzierten Topologie mit einer passenden, nur von z abhängigen Gewichtsfunktion $g(z) > 0$, für die $g(z) \cdot \sum_i f_i(y, z)$ μ -integrabel ist.

- c) Zu $q(y, z; u) \in Q$ gibt es eine Strategie $\tau_u(z)$ mit $q(y, z; u) = \vartheta(y, z; \tau_u(z))$.

Ferner sei S die Menge aller Funktionen $\delta(y, z; m)$, für die gilt:

- d) $s_i(y, z; m) f_i(y-m, z)$ ist μ -meßbar für alle i und alle $m \in R^1$.
- e) $\delta(y, z; m)$ ist darstellbar in der Gestalt $\delta(y, z; m) = \eta(y - m, z; v(y, z))$, wobei also $v(y, z)$ nicht von m abhängt.

Dann heißt S " vom Typ C (über $\mathcal{A}$) ".

Zur Verallgemeinerung des Begriffes der Bayeslösungen lassen wir uns von dem Gedanken einer gleichen Vorbewertung aller Parameterwerte m innerhalb jeder durch φ_i vorgewerteten Hypothese H_i leiten, vgl. auch E. Lehmann [47], Satz von Hunt - Stein. Mit der Bezeichnung

$r_{im}(\delta) := \int s_i(y, z; m) f_i(y-m, z) d\mu$ hätte man dann formal die Summe $\sum_i \varphi_i \int r_{im}(\delta) dm$ zu bilden, doch exi-

stieren die in der Summe auftretenden Integrale im allgemeinen nicht. Deshalb werden wir anstelle der Integration eine Mittelbildung über m benutzen. Ist dabei δ nur von $y-m$ und z abhängig, so wird $r_{im}(\delta)$ unabhängig von m und deshalb einfach mit $r_i(\delta)$ bezeichnet. Das tritt im Falle von (3.11) z.B. dann ein, wenn

$\delta(y, z; m) = \eta(y-m, z; v(z))$ ist bei passendem $v(z)$.

(3.12) Definition. S sei vom Typ C. Die Funktion $\delta^v(y, z; m) \in S$ heißt " Bayeslösung zur Vorbewertung φ " der zusammengesetzten Hypothesen H_i , wenn gilt:

- a) δ^v läßt sich darstellen in der Gestalt

$$\delta^v(y, z; m) = \eta(y-m, z; v_\varphi(z)) .$$

- b) Für alle $\delta \in S$ gilt

$$\sum_i \varphi_i r_i(\delta^v) \leq \liminf_{M=1,2,\dots} \frac{1}{2^M} \int \sum_{m=-M}^{+M} \varphi_i r_{im}(\delta) dm.$$

Schon wegen der Forderung (3.12 a) über die sehr spezielle Gestalt von $\delta\gamma$ ist die Existenz einer Bayeslösung nicht von vornherein sicher und wird unter bestimmten Voraussetzungen in Satz (3.15) bewiesen. Wegen

$$\liminf_{M=1,2,\dots} \frac{1}{2^M} \int_{m=-M}^{+M} \sum_i \varphi_i r_{im}(\delta) dm \leq \sup_m \sum_i \varphi_i r_{im}(\delta)$$

ist die Bayeslösung, falls sie existiert, zugleich eine gewisse Kombination aus Minimaxlösung bezüglich m und Bayeslösung zur Vorbewertung φ .

Später benötigen wir die beiden folgenden Hilfssätze.

(3.13) Hilfssatz. Es sei $\{g(y;u): u \in U\} =: G$ eine nicht leere Menge von Funktionen, die in $y \in R^1$ stückweise stetig sind mit höchstens abzählbar vielen, gemeinsamen Unstetigkeitsstellen. G sei kompakt bei der Abstandsdefinition

$$\Delta'(g(y;u_1), g(y;u_2)) = \sup_y \frac{1}{h(y)} |g(y;u_1) - g(y;u_2)|$$

mit einer Gewichtsfunktion $h(y) > 0$, für die $\int h(y) dy$ und $\int |y| h(y) dy$ existieren. Ferner mögen $\int |g(y;u)| dy$ und $\int |y g(y;u)| dy$ existieren für alle $u \in U$. Dann existiert für jede L -meßbare Funktion der Gestalt $g(y-m;v(y))$ das Integral $\int g(y-m;v(y)) dy$, und es gibt ein u_0 aus U derart, daß für alle L -meßbaren Funktionen $g(y-m;v(y))$ gilt

$$\liminf_{M=1,2,\dots} \frac{1}{2^M} \int_{m=-M}^{+M} \left[\int g(y-m;v(y)) dy \right] dm \geq \int g(y;u_0) dy.$$

Insbesondere ist $\int g(y;u_0) dy = \inf_{u \in U} \int g(y;u) dy$.

Beweis. Wegen der Kompaktheit von G ist

$\gamma(y) := \sup_{u \in U} |g(y; u)|$ L -meßbar, und analog zum Beweis von

Satz (3.3) folgt mit geeigneten $u_1, \dots, u_l \in U$

$$\int |y| \gamma(y) dy \leq \int |y| h(y) dy + \sum_{\lambda=1}^l \int |y| g(y; u_\lambda) dy$$

$$\leq C_1 < \infty$$

und ebenso

$$\int \gamma(y) dy \leq C_2 < \infty.$$

Da $g(y-m; v(y))$ in m stückweise stetig und bei festem m bezüglich y L -meßbar ist, ist $g(y-m; v(y))$ auch auf dem Produktraum (R_y^1, R_m^1) bezüglich des Produktmaßes meßbar *). Da außerdem

$$\begin{aligned} \int_{-M}^{+M} \left[\int_{-\infty}^{+\infty} |g(y-m; v(y))| dm \right] dy &\leq \int_{-M}^{+M} \left[\int_{-\infty}^{+\infty} \gamma(y-m) dm \right] dy \\ &\leq 2 M C_2 < \infty \end{aligned}$$

ist, ist im folgenden Beweis die Variablentransformation und die Vertauschung der Integrationsreihenfolge statthaft.

Nun ist

*) Sei nämlich $f(\xi, \eta)$ stetig in ξ und bei festem ξ L -meßbar bezüglich η . Dann sind die Funktionen

$$f_n(\xi, \eta) := f(v \cdot 2^{-n}, \eta) \quad \text{für} \quad v \cdot 2^{-n} \leq \xi < (v+1) 2^{-n},$$

$$v = \dots, -1, 0, 1, 2, \dots,$$

meßbar bezüglich des Produktmaßes, also auch der wegen der Stetigkeit existierende Limes

$$\lim_{n \rightarrow \infty} f_n(\xi, \eta) = f(\xi, \eta).$$

$$\begin{aligned}
& \int_{-M}^{+M} \left[\int_{-\infty}^{+\infty} g(y-m; v(y)) \, dy \right] \, dm = \int_{-\infty}^{+\infty} \left[\int_{p-M}^{p+M} g(p; v(y)) \, dy \right] \, dp \\
& = \int_{-\infty}^{+\infty} \left[\int_{-M}^{+M} g(p; v(y)) \, dy \right] \, dp + \int_{-\infty}^{+\infty} \left[\int_{p-M}^{-M} g(p; v(y)) \, dy \right] \, dp \\
& \quad + \int_{-\infty}^{+\infty} \left[\int_{+M}^{p+M} g(p; v(y)) \, dy \right] \, dp \\
& =: I_1 + I_2 + I_3 .
\end{aligned}$$

Dabei ist

$$\frac{1}{2M} I_1 \geq \inf_{u \in U} \int_{-\infty}^{+\infty} g(p; u) \, dp = \int_{-\infty}^{+\infty} g(p; u_0) \, dp ,$$

da das Infimum wegen der Kompaktheit von G angenommen wird. Weiterhin ist

$$\frac{1}{2M} |I_2| \leq \frac{1}{2M} \int_{-\infty}^{+\infty} |p| \, \gamma(p) \, dp \leq \frac{C_1}{2M} ,$$

$$\frac{1}{2M} |I_3| \leq \frac{C_1}{2M} .$$

Bei $M \rightarrow \infty$ folgt daraus die Behauptung des Hilfssatzes.

(3.14) Hilfssatz. $\mathfrak{K}'$ sei ein σ -Körper über M' , μ' ein vollständiges Maß über $\mathfrak{K}'$. H sei eine Menge von n -dimensionalen, reellen, μ' -meßbaren Funktionen $f(z)$ auf M' , $z \in M'$, und H sei kompakt bei der Abstandsdefinition

$$\Delta''(f, f^*) = \sup_z \sum_1 |h_1(z) - h_1^*(z)| .$$

$\mathcal{K}(z)$ sei eine beliebige, n -dimensionale, reelle, μ' -meßbare Funktion über M' . Dann ist

$K := \{z: \text{Es gibt ein } f(z) \in H \text{ mit } f(z) = \mathcal{K}(z)\}$
 μ' -meßbar.

Beweis. Zu jedem $c > 0$ gibt es endlich viele Funktionen $f^\lambda \in H$, $(\lambda=1,2,\dots,l)$ derart, daß für jedes $f \in H$ $\min_{\lambda \leq l} \Delta''(f, f^\lambda) \leq c$ gilt. Ferner ist

$$K(c) := \{z: \min_{\lambda \leq l} \sum_i |h_i^\lambda(z) - k_i(z)| \leq c\}$$

μ' -meßbar. Es konvergiere $c_1 > c_2 > \dots$ gegen Null.

Dann ist $K^* := \bigcap_{i=1}^{\infty} K(c_i)$ μ' -meßbar, und wie man leicht sieht, ist wegen der Kompaktheit von H schließlich $K^* = K$, w.z.b.w.

Wir zeigen nun, daß unter gewissen Voraussetzungen die in (3.12) definierten, verallgemeinerten Bayeslösungen existieren.

(3.15) Satz. S sei vom Typ C. Die Funktionen $q(y, z; u)$ aus Q seien stetig in y , die Dichten $f_i(y-m, z)$ stückweise stetig in y . Ferner mögen die Integrale

$$\int_{-\infty}^{+\infty} \sum_i |y f_i(y-m, z)| dy \quad \text{und}$$

$$\int_{-\infty}^{+\infty} \sum_i |y q_i(y-m, z; u) f_i(y-m, z)| dy \quad \text{für alle } u \in U$$

existieren, $m \in R^1$. Dann existiert zu jeder Vorbewertung φ der zusammengesetzten Hypothesen H_1 eine Bayeslösung δ^φ , und analog zu (3.10) gilt für μ' -fast alle $z \in M'$

$$\begin{aligned} \sum_i \varphi_i \int s_i^\varphi(y, z; m) f_i(y-m, z) dy \\ = \inf_{u \in U} \sum_i \varphi_i \int q_i(y-m, z; u) f_i(y-m, z) dy. \end{aligned}$$

Beweis. 1) Bei festem z setzen wir

$$g(y, z; u) := \sum_1 \varphi_1 q_1(y, z; u) f_1(y, z),$$

$$h(y, z) := \sum_1 f_1(y, z) + e^{-y^2} > 0.$$

Dann sind die Voraussetzungen von Hilfssatz (3.13) erfüllt, es existiert also ein $u_0(z)$ mit

$$\begin{aligned} \liminf_{M=1,2,\dots} \frac{1}{2^M} \int_{m=-M}^{+M} \left[\int \sum_1 \varphi_1 q_1(y-m, z; v(y, z)) f_1(y-m, z) dy \right] dm \\ \geq \int \sum_1 \varphi_1 q_1(y, z; u_0(z)) f_1(y, z) dy \\ = \inf_{u \in U} \int \sum_1 \varphi_1 q_1(y, z; u) f_1(y, z) dy \\ =: F(z). \end{aligned}$$

2) Wir zeigen als nächstes die Existenz einer Funktion $q(y, z; u_1(z))$, für die $\int q_1(y, z; u_1(z)) f_1(y, z) dy$ μ' -meßbar ist und die für μ' -fast alle $z \in M'$

$$(*) \quad \int \sum_1 \varphi_1 q_1(y, z; u_1(z)) f_1(y, z) dy = F(z)$$

erfüllt. Wir definieren

$$F_1(z) := \int f_1(y, z) dy,$$

$$p_1(z; u) := \begin{cases} \frac{1}{F_1(z)} \int q_1(y, z; u) f_1(y, z) dy & \text{für } F_1(z) \neq 0, \\ 0 & \text{für } F_1(z) = 0. \end{cases}$$

Dann besitzt $P := \{p(z; u): u \in U\}$ die in (3.2 a, b) geforderten Eigenschaften, wobei jetzt P statt Q , p statt q , z statt x steht und wir die Gewichtsfunktion $g(z)$ aus Definition (3.11) übernehmen; denn sei

$$d = \sup_{y, z} \left(\frac{1}{g(z)} \sum_1 |q_1(y, z; u_1) - q_1(y, z; u_2)| \right), \quad u_1, u_2 \in U,$$

dann gilt im Falle $F_1(z) \neq 0$

$$\begin{aligned} |p_1(z; u_1) - p_1(z; u_2)| \\ = \frac{1}{F_1(z)} \left| \int [q_1(y, z; u_1) - q_1(y, z; u_2)] f_1(y, z) dy \right| \\ \leq d \cdot g(z). \end{aligned}$$

Für $F_i(z) = 0$ ist auch $|p_i(z; u_1) - p_i(z; u_2)| = 0$, also gilt

$$\sup_z \left(\frac{1}{g(z)} \sum_i |p_i(z; u_1) - p_i(z; u_2)| \right) \leq n.d.,$$

so daß sich die Kompaktheit von Q auf P überträgt. - Da die Bedingung (3.2 c) beim Beweis der Sätze (3.9) und (3.10) nicht benötigt wird, folgt aus diesen Sätzen:

Es gibt eine Funktion $p(z; u_1(z))$ mit den Eigenschaften

a) $p_i(z; u_1(z)) F_i(z)$ ist μ' -integrierbar,

$$b) \sum_i \varphi_i p_i(z; u_1(z)) F_i(z) = \inf_{u \in U} \sum_i \varphi_i p_i(z; u) F_i(z)$$

für μ' -fast alle $z \in M'$.

Die letzte Gleichung ist aber nur eine andere Schreibweise für (*).

3) Wir zeigen schließlich noch: Wenn es eine Funktion $q(y, z; u_1(z))$ gibt, so daß

$$p_i(z; u_1(z)) F_i(z) = \int q_i(y, z; u_1(z)) f_i(y, z) dy$$

μ' -meßbar ist, so gibt es aufgrund der Kompaktheit von Q auch eine Funktion $q(y, z; u_2(z)) \in Q$ mit

$$\int q_i(y, z; u_2(z)) f_i(y, z) dy = p_i(z; u_1(z)) F_i(z) \quad \mu'\text{-fast}$$

derart, daß $q_i(y, z; u_2(z)) f_i(y, z)$ μ -meßbar ist.

Die Folge $c_1 > c_2 > \dots$ konvergiere gegen null. Zu c_1 gibt es eine Überdeckung von Q durch endlich viele, kompakte Teilmengen Q_{1j} ($1 \leq j \leq N_1$) derart, daß

$\bigcup_{j=1}^{N_1} Q_{1j} = Q$ gilt und der Durchmesser $d(Q_{1j}) \leq c_1$ ist für alle $j \leq N_1$. Weiterhin überdecken wir rekursiv $Q_{k-1, j}$

durch endlich viele, kompakte Teilmengen Q_{kl} ($n_{k, j-1} < 1 \leq n_{kj}$), so daß

$$\bigcup_{n_{k, j-1} < l \leq n_{kj}} Q_{kl} = Q_{k-1, j} \quad \text{und}$$

$d(Q_{kl}) \leq c_k$ gilt.

Für eine kompakte Teilmenge $Q' \subset Q$ setzen wir

$$A(Q') := \left\{ z: \quad \text{Es gibt ein } u \in U \text{ mit } \eta(y, z; u) \in Q' \text{ und} \right. \\ \left. \frac{1}{g(z)} p(z; u) = \frac{1}{g(z)} p(z; u_1(z)) \right\}.$$

Aus der Kompaktheit von Q' bezüglich des Abstandes Δ^* gemäß (3.11 b) folgt wie im Teil 2) des Beweises, daß

$$H := \left\{ \frac{1}{g(z)} p(z; u) : \quad \eta(y, z; u) \in Q' \right\}$$

kompakt ist bezüglich des Abstandes Δ'' aus Hilfssatz (3.14). Außerdem zieht aufgrund der Definition von $p(z; u)$ die μ' -Meßbarkeit von $p_i(z; u) F_i(z)$ und von $p_i(z; u_1(z)) F_i(z)$ auch die μ' -Meßbarkeit von $p_i(z; u)$ bzw. $p_i(z; u_1(z))$ nach sich, so daß wegen Hilfssatz (3.14) auch $A(Q')$ μ' -meßbar ist.

Wir weisen darauf hin, daß wir bei der Konstruktion der Q_{kl} den Abstand Δ^* der Funktionen $\eta(y, z; u)$ auf $M = (R^1, M')$ benutzten, dagegen bei der Definition von $A(Q')$ eine Gleichheitsrelation in M' . Dieser Unterschied ist für den Beweis wesentlich.

Sei nun jedem Q_{kl} eine Funktion $\eta(y, z; u_{kl}) \in Q_{kl}$ zugeordnet. Wir definieren

$$B_{kl} := A(Q_{kl}) \cap \left[\bigcap_{j=1}^{l-1} \overline{A(Q_{kj})} \right],$$

$$\delta^k := \eta(y, z; u_{kl}) \quad \text{auf} \quad (R^1, B_{kl}) \subset M.$$

Da $B_{kl} \subset M'$ μ' -meßbar ist, ist $s_i^k(y, z) f_i(y, z)$ μ -meßbar für alle i . Sofern

$$\lim_{k \rightarrow \infty} \delta^k(y, z) =: \delta^y(y, z) \quad \text{für } \mu'\text{-fast alle } z \in M'$$

existiert, ist dann auch $s_i^y(y, z) f_i(y, z)$ ($i=1, \dots, n$) μ -meßbar, und aufgrund der Konstruktion von $\delta^y(y, z)$ gilt mit einer geeigneten Funktion $u_2(z)$ für μ' -fast alle z $\delta^y(y, z) = \eta(y, z; u_2(z))$ und

$$\int q_1(y, z; u_2(z)) f_i(y, z) dy = p_i(z; u_1(z)) F_i(z).$$

Wir beweisen nun die Konvergenz der $\delta^k(y, z)$ für μ' -fast alle z . Aus

$$\bigcup_{n_{k,j-1} < l \leq n_{kj}} Q_{kl} = Q_{k-1,j}$$

folgt

$$\begin{aligned} \bigcup_{n_{k,j-1} < l \leq n_{kj}} B_{kl} &= \left[\bigcup_{n_{k,j-1} < l \leq n_{kj}} A(Q_{kl}) \right] \cap \left[\bigcap_{1 \leq l \leq n_{k,j-1}} \overline{A(Q_{kl})} \right] \\ &= A(Q_{k-1,j}) \cap \left[\bigcap_{l=1}^{j-1} \overline{A(Q_{k-1,l})} \right] \\ &= B_{k-1,j}. \end{aligned}$$

Da außerdem bei festem k alle B_{kl} disjunkt sind, gibt es zu jedem $B_{jm} \neq \emptyset$ und $k < j$ genau ein B_{kl} mit $B_{kl} \supset B_{jm}$, und es gilt $Q_{kl} \supset Q_{jm}$. Daraus ergibt sich

$$\Delta^*(\varphi(y, z; u_{kl}), \varphi(y, z; u_{jm})) \leq c_k,$$

und auf (R^1, B_{jm}) gilt

$$\begin{aligned} \sup \left(\frac{1}{g(z)} \sum_i |s_i^j(y, z) - s_i^k(y, z)| \right) \\ = \sup \left(\frac{1}{g(z)} \sum_i |q_i(y, z; u_{jm}) - q_i(y, z; u_{kl})| \right) \\ \leq c_k, \end{aligned}$$

wobei die Suprema über alle $(y, z) \in (R^1, B_{jm})$ zu nehmen sind. Daraus folgt die Konvergenz der Funktionen $\delta^k(y, z)$ für μ' -fast alle $z \in M'$.

4) Unter 1) bis 3) haben wir gezeigt: Zur Vorberwertung φ gibt es eine Funktion $\delta^q(y, z) = \varphi(y, z; u_2(z)) \in S$ derart, daß $s_i^q(y, z) f_i(y, z)$ μ -meßbar ist für alle i und daß für μ' -fast alle z und alle $\delta(y, z; m) \in S$ gilt:

$$\begin{aligned} \sum_i \varphi_i \int s_i^q(y, z) f_i(y, z) dy \\ \leq \liminf_{M=1,2,\dots} \frac{1}{2^M} \int_{m=-M}^{+M} \left[\int \sum_i \varphi_i s_i(y, z; m) f_i(y-m; z) dy \right] dm. \end{aligned}$$

Insbesondere genügt δ^q also der letzten Gleichung des Satzes (3.15). Durch Integration erhält man sofort

$$\begin{aligned} & \sum_i \varphi_i r_i(\delta^q) \\ & \leq \int \left\{ \liminf_{M=1,2,\dots} \frac{1}{2^M} \int_{m=-M}^{+M} \left[\int \sum_i \varphi_i s_i(y,z;m) f_i(y-m,z) dy \right] dm \right\} d\mu' \\ & \leq \liminf_{M=1,2,\dots} \frac{1}{2^M} \int_{m=-M}^{+M} \left[\sum_i \varphi_i \int s_i(y,z;m) f_i(y-m,z) d\mu \right] dm. \end{aligned}$$

δ^q ist also eine Bayeslösung zur Vorbewertung q . Damit ist Satz (3.15) bewiesen.

Die Sätze (3.13) bis (3.15) gelten sinngemäß auch noch, wenn man $y \in R^1$, $m \in R^1$ durch $(y_1, \dots, y_k) \in R^k$, $(m_1, \dots, m_k) \in R^k$ ersetzt.

Die Voraussetzungen des Satzes (3.15) bezüglich der Integrierbarkeit sind in den folgenden Paragraphen stets erfüllt, weil die Funktionen $q(y,z;u)$ nicht negativ sind und durch quadratische Polynome majorisiert werden und bei der Gaußverteilung alle Momente existieren. Die Kompaktheitsbedingungen sind in den Paragraphen 4 und 5 erfüllt, weil U dort nur endlich viele Elemente enthält; in § 6 wird die Kompaktheit gesondert gezeigt werden.

§ 4. Konstanter Schaden

Zunächst wollen wir die Ergebnisse der Theorie statistischer Entscheidungsfunktionen für den Fall näher untersuchen, daß der Schaden vom Ergebnis des Versuches nicht abhängt, sondern, wie in den Arbeiten über Hypothesentestung meistens angenommen wird, nur von der Hypothese und der Entscheidung.

Wir verwenden im übrigen die Bezeichnungen von § 3.

Unter den zusammengesetzten Hypothesen

H_i mit den Wahrscheinlichkeitsdichten $f_i(y-m, z)$
 $(i = 0, 1, \dots, n)$

seien die Entscheidungen

D_j : " H_j ist richtig " $(j = 0, 1, \dots, n)$
 zugelassen mit den Schäden $d_i(D_j) =: d_{ij}$.

Wir setzen in diesem und im nächsten Paragraphen

$\tau_u(z) = D_u \quad (u = 0, 1, \dots, n)$ und erhalten

$\eta(y, z; u) = \vartheta(y, z; D_u)$,

wobei in diesem Paragraphen außerdem $\vartheta(y, z; D_u) = \vartheta(D_u)$ von y und z unabhängig ist. Es ist also einfach

$Q = \{ \vartheta(D_j) : j=0, 1, \dots, n \}$.

Die Integrale auf der rechten Seite der letzten Gleichung von Satz (3.15) ergeben

$$\begin{aligned} \int d_{ij}(D_j) f_i(y-m, z) dy &= d_{ij} \int f_i(y, z) dy \\ &=: d_{ij} F_i(z). \end{aligned}$$

Die hierbei auftretenden Funktionen $F_i(z)$ sind gerade die Wahrscheinlichkeitsdichten unter den Hypothesen H_i über M' . Um eine Bayeslösung zu finden, hat man nach Satz (3.15) $j(z)$ so zu bestimmen, daß

$$(4.1) \quad \sum_{i=0}^n \varphi_i d_i(D_j(z)) F_i(z) = \min_j \sum_{i=0}^n \varphi_i d_{ij} F_i(z)$$

gilt für μ' -fast alle $z \in M'$.

Sei nun speziell

$$(4.2) \quad d_{ij} = \begin{cases} 0 & \text{für } i = j, \\ a & \text{für } i = 0, \quad j \neq 0, \\ b & \text{für } i \neq 0, \quad j = 0, \\ c & \text{sonst,} \end{cases} \quad a, b, c > 0,$$

$$\varphi_0 = 1 - n\varphi, \quad \varphi_1 = \dots = \varphi_n = \varphi, \quad 0 \leq \varphi \leq 1/n.$$

Diese Symmetrien erscheinen sinnvoll beim Ausreißerproblem mit höchstens einem Ausreißer, d.h. bei folgenden Hypothesen über n Variable $a_1, \dots, a_n$:

Unter H_0 sind alle n Variablen nach $\psi_0(x)$ verteilt; unter H_i ($i=1, \dots, n$) ist a_i nach $\psi_1(x)$ verteilt und a_j ($j \neq i$) nach $\psi_0(x)$.

Bei diesen Voraussetzungen über die Konstanten d_{ij} und φ_i liefert die Bayeslösung (4.1) die Entscheidung D_0 , falls

$$a(1 - n\varphi) F_0(z) + \left[(c - b) \sum_{i=1}^n F_i(z) - c F_j(z) \right] \varphi \geq 0$$

für $j = 1, \dots, n$;

ist diese Bedingung nicht erfüllt, so erhält man bei der Bayeslösung diejenige Entscheidung D_j ($j > 0$), für die $F_j(z)$ den größten Wert annimmt. Im Falle $F_0(z) \neq 0$, $\varphi \neq 0$ kann man die obige Ungleichung in der Gestalt

$$(4.3) \quad a \frac{1 - n\varphi}{\varphi} + (c - b) \sum_{i=1}^n \frac{F_i(z)}{F_0(z)} - c \frac{F_j(z)}{F_0(z)} \geq 0$$

für $j = 1, \dots, n$

schreiben.

Interessant ist der Spezialfall $a = b = c$. Man

gelangt dann für $F_0(z) \neq 0$, $\varphi \neq 0$ zu einer Bayeslösung, wenn man die Entscheidung

$$(4.4) \quad D_0 \text{ bei } C_0 = \frac{1 - n\varphi}{\varphi} \geq \frac{F_j(z)}{F_0(z)} \quad \text{für } j=1, \dots, n,$$

$$D_j \text{ bei } C_0 = \frac{1 - n\varphi}{\varphi} < \frac{F_j(z)}{F_0(z)} \quad \text{und}$$

$$F_j(z) = \max_{k \geq 0} F_k(z)$$

trifft. Damit erhalten wir aber gerade den t & k -optimalen Test von J. Pfanzagl [32], der seinerseits die Tests von E. Paulson, D. R. Truax usw. umfaßt, vgl. § 2. Statt durch φ wird C_0 bei den genannten Autoren jedoch durch die vorgegebene Größe des Fehlers erster Art bestimmt.

Der in diesem Spezialfall der Rechnung zugrunde gelegte Schaden

$d_{ij} = 0$, wenn die richtige Entscheidung gewählt wird,
 $d_{ij} = a$, wenn eine falsche Entscheidung gewählt wird,

berücksichtigt aber überhaupt nicht, wie sehr die fälschlich weggelassene bzw. die fälschlich nicht weggelassene Beobachtung die weiteren Ergebnisse (z.B. den Schätzwert für den Erwartungswert) beeinflusst. Deshalb werden wir im nächsten Paragraphen eine andere Schadensfunktion einführen.

Beispiel. Die Hypothesen über n Stichproben ungleichen Umfanges N_v ($v=1, \dots, n$) von unabhängigen Gaußvariablen $\xi_{v\mu}$ ($v=1, \dots, n$; $\mu=1, \dots, N_v$) mit dem unbekannten, gemeinsamen Erwartungswert m und der ebenfalls unbekannten Varianz σ_i^2 der i -ten Stichprobe seien

H_0 mit $\sigma_1^2 = \dots = \sigma_n^2 =: \sigma^2 > 0$, $m \in \mathbb{R}^1$,

H_i mit $\sigma_i^2 = k^2 \sigma^2$, $\sigma_j^2 = \sigma^2 > 0$ ($j \neq i$), $m \in \mathbb{R}^1$ ($i=1, \dots, n$)

bei bekanntem $k > 1$ und unbekanntem m und σ .

Wir suchen die Bayeslösungen. Dazu setzen wir

$$N := \sum_{v=1}^n N_v, \quad K := 1 - k^{-2}$$

und führen als neue Variable ein

$$(4.5) \quad \bar{x} := \frac{1}{N} \sum_{v,\mu} x_{v\mu}, \quad s := +\sqrt{\frac{1}{N-1} \sum_{v,\mu} (x_{v\mu} - \bar{x})^2},$$

$$t_{v\mu} := \frac{x_{v\mu} - \bar{x}}{s}.$$

Wegen $\sum_{v,\mu} t_{v\mu} = 0$ und $\sum_{v,\mu} t_{v\mu}^2 = N - 1$ lassen sich zwei der $t_{v\mu}$ durch die restlichen ausdrücken; bezeichnen wir sie mit t' und t'' und bedeutet $\sum^*$ die Summation über alle $t_{v\mu}$ außer t' und t'' , so gilt

$$(4.6) \quad t', t'' = -\frac{1}{2} \sum^* t_{v\mu} \pm \sqrt{\frac{N-1}{2} - \frac{1}{2} \sum^* t_{v\mu}^2 - \frac{1}{4} \left(\sum^* t_{v\mu} \right)^2}.$$

In den folgenden Gleichungen sind t' und t'' vermöge (4.6) als Funktionen der restlichen $t_{v\mu}$ aufzufassen.

Die Wahrscheinlichkeitsdichten $f_1(\bar{x}, s, t_{v\mu})$ verschwinden außerhalb des Bereiches

$$\left\{ \sum^* t_{v\mu}^2 + \frac{1}{2} \left(\sum^* t_{v\mu} \right)^2 \leq N - 1 \right\} \cap \{s \geq 0\};$$

innerhalb dieses Bereiches gilt für $i \neq 0$:

$$f_1(\bar{x}, s, t_{v\mu}) = (2\pi\sigma^2)^{-N/2} k^{-N_1} \exp \left\{ -\frac{s^2}{2\sigma^2} \left[(N-1) + \frac{N(\bar{x} - m)^2}{s^2} - K N_1 \frac{(\bar{x} - m)^2}{s^2} - 2K \frac{\bar{x} - m}{s} \sum_{\mu=1}^{N_1} t_{i\mu} - K \sum_{\mu=1}^{N_1} t_{i\mu}^2 \right] \right\} \frac{2N(N-1)s^{N-2}}{|t' - t''|}.$$

Hieraus erhält man f_0 , indem man $k=1$ ($K=0$) setzt:

$$f_0(\bar{x}, s, t_{v\mu}) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{s^2}{2\sigma^2} (N-1) + \frac{N(\bar{x} - m)^2}{s^2} \right\} \frac{2N(N-1)s^{N-2}}{|t' - t''|}.$$

Führt man statt $\bar{x}$ und s nun $\bar{x}' := \bar{x}/\sigma$ und $s' := \log s$ ein, so treten die unbekannten Parameter m und σ nur mehr in der Verbindung $\bar{x}' - m/\sigma$ bzw. $s' - \log \sigma$ auf, wie es in § 3 ab (3.11) vorausgesetzt wird. Wir haben hier zwei Variable $\bar{x}'$ und s' statt der Variablen y in § 3 und entsprechend m/σ und $\log \sigma$ statt m , doch gelten die Ausführungen des § 3 sinngemäß auch in diesem Falle. Die Bayeslösungen (3.12) verlangen hier also eine Mittelbildung des Risikos über m/σ und $\log \sigma$.

Die Integration über $\bar{x}$ und s , die natürlich dasselbe liefert wie die Integration der transformierten Dichten über $\bar{x}'$ und s' , ergibt für $i \neq 0$

$$\begin{aligned} F_i(t_{v\mu}) &:= \int f_i(\bar{x}, s, t_{v\mu}) d\bar{x} ds \\ &= (2\pi\sigma^2)^{-(N-1)/2} \int k^{-N_i} \exp \left\{ -\frac{1}{2} \frac{s^2}{\sigma^2} \left[N-1 - K \sum_{\mu=1}^{N_i} t_{i\mu}^2 \right. \right. \\ &\quad \left. \left. - \frac{K^2}{N - N_i K} \left(\sum_{\mu=1}^{N_i} t_{i\mu} \right)^2 \right] \right\} \frac{2N(N-1)s^{N-2}}{|t' - t''| \sqrt{N - N_i K}} ds \\ &= \frac{\pi^{-(N-1)/2} \Gamma\left(\frac{N-1}{2}\right) N(N-1)}{|t' - t''| k^{N_i} \sqrt{N - N_i K}} \left[(N-1) \right. \\ &\quad \left. - K \sum_{\mu=1}^{N_i} t_{i\mu}^2 - \frac{K^2}{N - N_i K} \left(\sum_{\mu=1}^{N_i} t_{i\mu} \right)^2 \right]^{-\frac{N-1}{2}}, \\ (4.7) \quad \frac{p_i(t_{v\mu})}{F_0(t_{v\mu})} &= k^{-N_i} \sqrt{\frac{N}{N - N_i K}} \left[1 - \frac{K}{N-1} \sum_{\mu=1}^{N_i} t_{i\mu}^2 \right. \\ &\quad \left. - \frac{K^2}{(N-1)(N - N_i K)} \left(\sum_{\mu=1}^{N_i} t_{i\mu} \right)^2 \right]^{-\frac{N-1}{2}}. \end{aligned}$$

Ist speziell $N_i = 1$ für alle i , so erhält man mit

$$A := \frac{1}{K} \sqrt{\frac{N}{N-K}} \quad \text{und} \quad B := \frac{N}{N-1} \frac{K}{N-K}$$

$$(4.8) \quad \frac{F_i(t_v)}{F_0(t_v)} = A \left[1 - B t_i^2 \right]^{-\frac{N-1}{2}}.$$

Die Ergebnisse (4.7) bzw. (4.8) sind nun in (4.3) und (4.4) einzusetzen, um eine Bayeslösung des Ausreißerproblems unter den angegebenen Hypothesen und für den Schaden (4.2) zu erhalten.

Wir betrachten noch einmal den Spezialfall $a = b = c$. Für $N_1 = \dots = N_n = 1$ lassen sich dann die Bayeslösungen in die folgende Form bringen:

Man treffe die Entscheidung

$$\begin{aligned} D_0 & \quad \text{bei} \quad t_i^2 \leq C(\varphi, k, n) \quad \text{für alle} \quad i=1, \dots, n, \\ D_j & \quad \text{bei} \quad t_j^2 > C(\varphi, k, n) \quad \text{und} \quad t_j^2 = \max_v t_v^2, \quad j=1, \dots, n. \end{aligned}$$

Wie man zeigen kann, ist die Gesamtheit dieser Bayeslösungen für $0 \leq \varphi \leq 1/n$ identisch mit der Gesamtheit aller Entscheidungsverfahren mit der Eigenschaft (2.1) für $0 \leq \alpha \leq 1$, wobei dort lediglich die Konstante C durch α und n zu bestimmen ist statt durch φ , n und k bei unserer Interpretation als Bayeslösungen.

Im allgemeineren Falle $N_1 = \dots = N_n =: L \neq 1$ lauten bei den Bayeslösungen die Entscheidungen jedoch

$$\begin{aligned} D_0 & \quad \text{für} \quad T_i := \sum_{\mu=1}^L t_{i\mu}^2 + \frac{K}{N-LK} \left(\sum_{\mu=1}^L t_{i\mu} \right)^2 \\ & \leq C(\varphi, k, n, L) \quad \text{für alle} \quad i = 1, \dots, n, \\ D_j & \quad \text{für} \quad T_j > C(\varphi, k, n, L) \quad \text{und} \quad T_j = \max_v T_v, \quad j=1, \dots, n. \end{aligned}$$

Die linke Seite dieser Ungleichungen hängt von k ab, und auch durch weitere Umformungen kann man nicht die

Gestalt $h(t_{i\mu}; n, L) \left\{ \begin{array}{l} \leq \\ > \end{array} \right\} C^*(\varphi, k, n, L)$ erreichen mit passenden Funktionen h und C^* . Deshalb gibt es, wie man zeigen kann, für die hier betrachteten Hypothesen keinen von k unabhängigen Test mit der Eigenschaft (2.1) im Gegensatz zu den Hypothesen, die E. Paulson, D. R. Truax u. a. benutzten, vgl. § 2.

§ 5. Quadratischer Schaden und endlich viele Entscheidungen

Im allgemeinen will man nicht nur wissen, ob ein Ausreißer vorhanden ist, sondern man möchte anschließend die unbekannten Parameter schätzen. Wäre etwa bekannt, daß die Gaußvariable g_1 die Varianz σ^2 besitzt, die Gaußvariablen $g_2, \dots, g_n$ aber die Varianz 1 und alle g_i den gleichen, unbekannten Erwartungswert m , so wäre für diesen die beste Statistik, d.h. die mit der kleinsten Varianz, die Größe

$$b_1 := \frac{1}{n-1 + \sigma^{-2}} \left(g_1 \sigma^{-2} + \sum_{v=2}^n g_v \right).$$

Trifft man beim Ausreißertest die Entscheidung, daß g_1 ein Ausreißer ist, so ist es deshalb sinnvoll, b_1 als Statistik für m zu benutzen.

Daß die Statistiken für m , die wir in diesem und im folgenden Paragraphen erhalten, auch die Ausreißer berücksichtigen, liegt letzten Endes daran, daß die den Ausreißer kennzeichnenden Parameter (also etwa der Quotient der Varianzen) als bekannt vorausgesetzt werden.

Ist über die Verteilung eines etwa vorhandenen Ausreißers gar nichts bekannt, so kann man nur die außerhalb einer gewissen Schranke liegenden Beobachtungen weglassen, mit den übrigen die gesuchten Parameter schätzen und hoffen, ein annähernd richtiges Ergebnis zu erhalten. Weiß man etwa, daß der Ausreißer nur einen anderen Erwartungswert oder nur eine andere Varianz besitzt als die übrigen Variablen, so wird durch die von E.Paulson, D.R.Truxax, K.V.Ramachandran und C.G.Khatri, R.Doornbos und H.Prins, A.Kudo sowie J.Pfanzagl angegebenen Testverfahren zu vorgeschriebenem Fehler erster Art der Fehler zweiter Art zum Minimum gemacht, und zwar unabhängig

von der Größe des Ausreißerparameters, solange alle Stichproben den gleichen Umfang haben. Wie groß jedoch der Fehler zweiter Art ist, hängt natürlich vom Ausreißerparameter ab. Ist dieser aber bekannt, so kann der Ausreißer auch mit zur Parameterschätzung verwendet werden; andernfalls verzichtet man auf einen Teil der Information, die das Experiment geliefert hat. In der Praxis wird der Ausreißerparameter allerdings selten exakt bekannt sein; doch kennt man häufig wenigstens seine Größenordnung.

Für die unabhängigen Gaußvariablen $g_1, \dots, g_n$ mit der Varianz σ_v^2 ($v=1, \dots, n$) und dem gemeinsamen, unbekannten Erwartungswert m betrachten wir jetzt die zusammengesetzten Hypothesen

$$\begin{aligned} H_0 & \text{ mit } \sigma_v^2 = 1 \quad \text{für } v = 1, \dots, n, \quad m \in R^1, \\ H_1 & \text{ mit } \sigma_1^2 = \sigma^2 > 1, \quad \sigma_v^2 = 1 \quad \text{für } v \neq 1, \quad m \in R^1 \\ & \quad \quad \quad (i = 1, \dots, n) \end{aligned}$$

bei bekanntem σ^2 . Wir wollen m schätzen und benutzen, wenn $h_j(x)$ die bei der Entscheidung D_j gewählte Schätzfunktion ist, den Schaden

$$(5.1) \quad d_1(x; D_j) = (h_j(x) - m)^2,$$

der also unabhängig von H_1 ist.

Wir führen anstelle von $x_1, \dots, x_n$ neue Variable ein:

$$(5.2) \quad \bar{x} := \frac{1}{n} \sum_{v=1}^n x_v, \quad z_v := x_v - \bar{x}.$$

Wegen $\sum z_v = 0$ ist wie im letzten Paragraphen eine der Variablen z_v , z.B. z_n , überflüssig; sie wird aber beibehalten, um die Formeln zu vereinfachen. Weiterhin führen wir die folgenden Abkürzungen ein:

$$(5.3) \quad k := (n-1 + \sigma^{-2}) \sigma^2, \quad K := \frac{n-1 + \sigma^{-2}}{n(1 - \sigma^{-2})}.$$

Die zulässigen Entscheidungen und die zugehörigen Schätzfunktionen $h_j(\mathbf{x})$ seien in Übereinstimmung mit den am Anfang dieses Paragraphen gemachten Ausführungen:

$$(5.4) \quad \begin{aligned} D_0 & \quad \text{"kein Ausreißer"}, & h_0(\mathbf{x}) &= \bar{x}; \\ D_j & \quad \text{"} g_j \text{ ist ein Ausreißer"}, \\ & h_j(\mathbf{x}) = \frac{1}{n-1 + \sigma^{-2}} \left(x_j \sigma^{-2} + \sum_{v \neq j} x_v \right) \\ & = \bar{x} - \frac{1}{nK} z_j \quad (j=1, \dots, n). \end{aligned}$$

Die Wahrscheinlichkeitsdichten $f_i(\bar{x}, z_1, \dots, z_{n-1}; m)$ lauten wegen $\frac{\partial(x_1, \dots, x_n)}{\partial(\bar{x}, z_1, \dots, z_{n-1})} = n$

$$\begin{aligned} f_i(\bar{x}, z_1, \dots, z_{n-1}; m) \\ = \frac{(2\pi)^{-n/2}}{\sigma} n \exp \left\{ -\frac{1}{2} (\bar{x} - m)^2 k \sigma^{-2} - \frac{1}{2} \sum_v z_v^2 \right. \\ \left. + (1 - \sigma^{-2}) z_i (\bar{x} - m) + \frac{1}{2} (1 - \sigma^{-2}) z_i^2 \right\} \quad (i=1, \dots, n), \end{aligned}$$

$$\begin{aligned} f_0(\bar{x}, z_1, \dots, z_{n-1}; m) \\ = n (2\pi)^{-n/2} \exp \left\{ -\frac{n}{2} (\bar{x} - m)^2 - \frac{1}{2} \sum_v z_v^2 \right\}. \end{aligned}$$

Um gemäß Satz (3.15) die Bayeslösungen zu gewinnen, benötigen wir die Integrale

$$S_{ij}(z) := \int d_i(\bar{x}, z; D_j) f_i(\bar{x}, z) d\bar{x}.$$

Mit $c := n(2\pi)^{-(n-1)/2}$ gilt für $i \neq 0, j \neq 0$

$$S_{ij}(z) = \frac{c}{\sqrt{k}} \left[\frac{\sigma^2}{k} + \frac{(z_i - z_j)^2}{n^2 K^2} \right] \exp \left\{ -\frac{1}{2} \sum_v z_v^2 + \frac{1}{2K} z_i^2 \right\}.$$

Weiterhin ist

$$S_{0j}(z) = \frac{c}{\sqrt{n}} \left[\frac{1}{n} + \frac{z_j^2}{n^2 K^2} \right] \exp \left\{ -\frac{1}{2} \sum z_v^2 \right\} \quad (j \neq 0),$$

$$S_{i0}(z) = \frac{c}{\sqrt{k}} \left[\frac{\sigma^2}{k} + \frac{z_1^2}{n^2 K^2} \right] \exp \left\{ -\frac{1}{2} \sum z_v^2 + \frac{1}{2K} z_1^2 \right\} \quad (i \neq 0),$$

$$S_{00}(z) = \frac{c}{\sqrt{n}} \left[\frac{1}{n} + 0 \right] \exp \left\{ -\frac{1}{2} \sum z_v^2 \right\}.$$

Die beiden Summanden von $S_{ij}(z)$, die den beiden in den eckigen Klammern stehenden Gliedern entsprechen, bezeichnen wir mit $S'_{ij}(z)$ und $S''_{ij}(z)$. $S'_{ij}(z)$ hängt von der getroffenen Entscheidung nicht ab und spielt daher bei der Berechnung der Bayeslösungen keine Rolle; bei der Berechnung der Minimaxlösung muß $S'_{ij}(z)$ jedoch berücksichtigt werden. $S'_{ij}(z)$ gibt den über m gemittelten Schaden an, der bei der richtigen Entscheidung eintritt, d.h. wenn man unter der Hypothese H_1 die Entscheidung D_1 trifft. Geht man bei gleicher Bedeutung der $h_j(x)$ von dem Schaden

$$(5.5) \quad d_i^*(x; D_j) = [h_1(x) - h_j(x)]^2$$

aus statt von (5.1), so erhält man von vornherein $S''_{ij}(z)$ statt $S_{ij}(z)$.

Die Bayeslösungen zu der symmetrischen Vorbewertung $(1 - n\varphi, \varphi, \dots, \varphi)$ der zusammengesetzten Hypothesen $H_0, H_1, \dots, H_n$ erhält man aufgrund des Satzes (3.15) durch Aufsuchen des kleinsten der zu $\sum_i \varphi_i S''_{ij}$ proportionalen Ausdrücke

$$\varphi \sqrt{\frac{n}{k}} \sum_i z_i^2 \exp \left\{ \frac{1}{2K} z_1^2 \right\} \quad \text{für } j = 0 \quad \text{und}$$

$$(1-n\varphi) z_j^2 + \varphi \sqrt{\frac{n}{k}} \sum_i (z_i - z_j)^2 \exp \left\{ \frac{1}{2K} z_1^2 \right\} \quad \text{für } j=1, \dots, n.$$

Der Test schreibt also vor, g_j als Ausreißer zu behandeln, wenn

$$\min_v \left[\sqrt{\frac{k}{n}} \frac{1 - n\varphi}{\varphi} z_v^2 + z_v \sum_i (z_v - 2z_i) \exp \left\{ \frac{z_i^2}{2K} \right\} \right] < 0$$

ist und das Minimum bei $v = j$ eintritt.

Der Bereich, in dem eine Entscheidung D_j angenommen wird, läßt sich nicht explizit in der Gestalt

$$|z_j| > F_{\varphi}(z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_n)$$

angeben; ist $\exp \{ z_j^2 / 2K \} \gg \exp \{ z_v^2 / 2K \}$ für alle $v \neq j$, so ist näherungsweise g_j dann als Ausreißer zu behandeln,

wenn $z_j^2 > 2K \log \left(\sqrt{\frac{k}{n}} \frac{1 - n\varphi}{\varphi} \right)$ ist.

Dieser Test besitzt eine auf den ersten Augenblick überraschend wirkende Eigenschaft: Wenn die Entscheidung D_j ($j \neq 0$) angenommen wird, so braucht x_j weder die größte noch die kleinste Beobachtung zu sein, sondern kann irgendwo in der Mitte liegen. Daß das so sein muß, kann man sich wie folgt plausibel machen:

Bei der Wahl der Schadensfunktionen (5.1) oder (5.5) suchen wir eigentlich nicht einen Ausreißer, sondern die beste Schätzfunktion für den Erwartungswert. Liegen alle Beobachtungen beieinander, so ergibt der Test, wie man erwartet, $h_0(x) = \frac{1}{n} \sum x_v$, wobei also alle Beobachtungen das gleiche Gewicht $1/n$ haben. Liegt x_j weit außerhalb der übrigen Beobachtungen, so ergibt der Test, wie man wiederum erwartet, $h_j(x) = (x_j \sigma^{-2} + \sum_{v \neq j} x_v) / (k \sigma^{-2})$, wobei also x_j nur das σ^{-2} -fache Gewicht der übrigen Beobachtungen besitzt. Liegt nun x_j zwar außerhalb der übrigen Beobachtungen, aber nicht zu weit, so weiß man nicht recht, ob man x_j als Ausreißer behandeln soll oder nicht, und es scheint vernünftig, x_j ein Gewicht zu geben, das zwischen den Extremen bei $h_0(x)$ bzw. $h_j(x)$

liegt. Das tut der Test aber gerade: Er ergibt weder

$$h_0(x) = \bar{x} \quad \text{noch} \quad h_j(x) = \bar{x} - \frac{1}{n \cdot K} z_j, \quad \text{sondern}$$

$$h_i(x) = \bar{x} - \frac{1}{n \cdot K} z_i \quad \text{mit einem geeigneten } z_i \text{ zwischen } 0 \text{ und } z_j \text{ anstelle der hier nicht zugelassenen Schätzfunktion } \bar{x} - \lambda \frac{1}{n \cdot K} z_j \quad \text{mit } 0 < \lambda < 1.$$

Die Entscheidung D_j dürfte deshalb eigentlich nicht "g_j ist ein Ausreißer" genannt werden, sondern müßte "h_j ist die beste Schätzfunktion" heißen.

Es liegt nun nahe, auch allgemeinere Schätzfunktionen zuzulassen. Das wird im folgenden Paragraphen geschehen.

§ 6. Quadratischer Schaden

und unendlich viele Entscheidungen

Wir gehen von den gleichen Hypothesen über n unabhängige Gaußvariable mit gemeinsamem Erwartungswert m aus wie im letzten Paragraphen, nämlich von:

H_0 mit $\sigma_1^2 = \dots = \sigma_n^2 = 1$, $m \in R^1$;

H_i mit $\sigma_i^2 = \sigma^2 > 1$, $\sigma_j^2 = 1$ ($j \neq i$), $m \in R^1$ ($i=1, \dots, n$).

Dabei wird σ^2 als bekannt vorausgesetzt.

Wir benutzen weiterhin die in (5.2) bis (5.4) eingeführten Bezeichnungen $\bar{x}$, z_i , k , K und $h_j(z)$.

Der Raum T der Terminalentscheidungen sei eine Gesamtheit D_a mit $-\infty < a < +\infty$, wobei D_a vorschreibe, daß beim Meßergebnis x die Differenz $\bar{x} - a$ als Schätzwert für das unbekannte m zu nehmen ist. Eine Entscheidungsfunktion ist dann eine meßbare, reelle Funktion $a(\bar{x}, z)$, die dem gemessenen x für m den Schätzwert $\bar{x} - a(\bar{x}, z)$ zuordnet.

Analog zu (5.5) benutzen wir bei der Hypothese H_i und der Entscheidung D_a als Schaden $d_i(\bar{x}, z; m; D_a)$ den Ausdruck $[h_i - (\bar{x} - a)]^2$ und erhalten

$$(6.1) \quad d_0(\bar{x}, z; m; D_a) = a^2,$$

$$d_i(\bar{x}, z; m; D_a) = \left[a - \frac{1}{n \cdot K} z_i \right]^2 \quad (i=1, \dots, n).$$

Da in d_i weder $\bar{x}$ noch m auftritt, ist die vor (3.11) gestellte Forderung trivialerweise erfüllt, daß der Schaden nicht von $\bar{x}$ und m einzeln, sondern nur von $\bar{x} - m$ abhängen soll.

Wir schreiben von nun an $d_i(z; a)$ statt $d_i(\bar{x}, z; m; D_a)$.

Um (3.11) bis (3.15) anwenden zu dürfen, können wir für Q nicht die Gesamtheit aller Funktionen $\vartheta(z; a)$ nehmen, weil dann Q nicht kompakt wäre. Nun wird aber für einen optimalen Schätzwert $\bar{x} - a(\bar{x}, z)$ sicherlich $\min_i x_i \leq \bar{x} - a(\bar{x}, z) \leq \max_i x_i$ gelten, also erst recht $|a(\bar{x}, z)| \leq 1 + \max_i |z_i| =: G(z)$. Deshalb setzen wir mit festem $A \geq 1$

$$\vartheta(z; u) := \vartheta(z; u \cdot G(z)),$$

$$Q := \{ \vartheta(z; u) : |u| \leq A \}.$$

Q ist bei der Wahl $g(z) := G(z)^2$ kompakt; denn für $i \neq 0$ gilt

$$\begin{aligned} \sup_z \left(\frac{1}{g(z)} |q_i(z; u_1) - q_i(z; u_2)| \right) \\ = \sup_z |u_1 - u_2| \cdot \left| u_1 + u_2 - \frac{2}{n \cdot K \cdot G(z)} z_i \right| \\ \leq |u_1 - u_2| \left(2A + \frac{2}{n \cdot K} \right); \end{aligned}$$

weiterhin ist

$$\begin{aligned} \sup_z \left(\frac{1}{g(z)} |q_0(z; u_1) - q_0(z; u_2)| \right) \\ = |u_1^2 - u_2^2| \leq 2A |u_1 - u_2|, \end{aligned}$$

und daraus folgt sofort die Kompaktheit von Q .

Um die Bayeslösungen unter allen meßbaren Schätzfunktionen $\bar{x} - a(\bar{x}, z)$ mit der Einschränkung $|a(\bar{x}, z)| \leq A \cdot G(z)$ zu finden, sind gemäß Satz (3.15) die Integrale $S_{iu}^*(z) := \int q_i(z; u) f_i(\bar{x}, z) d\bar{x}$ zu bilden, und es ist das Infimum von $\sum \varphi_i S_{iu}^*(z)$, $|u| \leq A$, zu suchen. Mit der Bezeichnung

$$S_{ia}(z) := \int d_i(z; a) f_i(\bar{x}, z) d\bar{x}$$

gilt aber für festes z

$$\inf_{|u| \leq A} \sum \varphi_i S_{iu}^*(z) = \inf_{|a| \leq AG(z)} \sum \varphi_i S_{ia}(z).$$

Die Bayeslösungen zur Vorbewertung $(1-n\varphi, \varphi, \dots, \varphi)$ der zusammengesetzten Hypothesen $H_0, H_1, \dots, H_n$ erhalten wir deshalb aus dem Minimum über alle a unter der Nebenbedingung $|a| \leq A G(z)$ von $\sum_{i=0}^n \varphi_i S_{ia}(z)$.

Ähnlich wie in § 5 gilt mit $c := n(2\pi)^{-(n-1)/2}$

$$S_{ia}(z) = \frac{c}{\sqrt{K}} \left[a - \frac{z_i}{nK} \right]^2 \exp \left\{ -\frac{1}{2} \sum z_v^2 + \frac{1}{nK} z_i^2 \right\} \quad (i \neq 0),$$

$$S_{0a}(z) = \frac{c}{\sqrt{n}} \cdot a^2 \exp \left\{ -\frac{1}{2} \sum z_v^2 \right\}.$$

Mit der Abkürzung

$$(6.2) \quad C := \sqrt{\frac{K}{n}} \frac{1 - n\varphi}{\varphi}$$

ergibt sich für die Bayeslösung

$$(6.3) \quad a(z) = \frac{\sum z_i \exp \{ z_i^2 / 2K \}}{nK (C + \sum \exp \{ z_i^2 / 2K \})}.$$

Wie man leicht sieht, ist $|a(z)| \leq \max |z_v|/nK$, also auch $|a(z)| \leq G(z)$; die Bayeslösung ist somit unabhängig von der Einschränkung $|u| \leq A$, da $A \geq 1$ vorgeschrieben ist. Die genannte Einschränkung hat sich damit nachträglich als unwesentlich herausgestellt und ist insofern gerechtfertigt worden.

Die Abhängigkeit der Funktion $a(z)$ von φ steckt nur in C . Ist $\varphi = 0$, mißt man also den Hypothesen, die das Vorhandensein eines Ausreißers annehmen, keine subjektive Wahrscheinlichkeit zu, so wird $a(z) \equiv 0$. Die Korrektur $a(z)$ zu $\bar{x}$ wird immer größer, je größer φ ist. Für $\varphi = 1/n$ wird $C = 0$.

Ist $|z_i| \gg |z_j|$ für alle $j \neq i$ und $|z_i| \gg C$, so ist

$a(z) \approx z_1/nK$, und man erhält als Schätzfunktion für den Mittelwert näherungsweise die Funktion $h_1(z)$ des letzten Paragraphen. Je kleiner $|z_1|$ ist, desto kleiner ist auch der Faktor von z_1 in (6.3), wie man es nach den Überlegungen am Ende des letzten Paragraphen erwartet. Der Test liefert im Gegensatz zu § 5 auch formal nicht mehr eine Aussage der Gestalt " g_j ist ein Ausreißer".

Die Risiken r'_0 bzw. r'_1 ($i \neq 0$) unter den Hypothesen H_0 bzw. H_1 ($i \neq 0$) betragen für die Bayeslösung (6.3) mit der Abkürzung dz für $dz_1 \dots dz_{n-1}$

$$(6.4) \quad r'_i = \frac{c}{\sqrt{K}} \int \left[a(z) - \frac{1}{nK} z_i \right]^2 \exp \left\{ -\frac{1}{2} \sum_{v=1}^n z_v^2 + \frac{z_i^2}{2K} \right\} dz$$

$$(i = 1, \dots, n),$$

$$r'_0 = \frac{c}{\sqrt{n}} \int a(z)^2 \exp \left\{ -\frac{1}{2} \sum_{v=1}^n z_v^2 \right\} dz,$$

wobei wieder z_n für $-\sum_{v=1}^{n-1} z_v$ steht. Diese Integrale lassen sich schon für $n = 3$ nicht mehr geschlossen auswerten (für $n = 2$ ist $a(z) \equiv 0$).

Mit den Bezeichnungen z_M und $\mathcal{L}$ gemäß

$$(6.5) \quad |z_M| := \max_v |z_v|,$$

$$\mathcal{L} := \{(\bar{x}, z): |z_M| \leq B\}$$

wird häufig die folgende, wesentlich einfachere Schätzfunktion für den Erwartungswert benutzt unter der Voraussetzung, daß höchstens ein Ausreißer vorkommt:

$$(6.6) \quad \bar{x} - b(z) := \begin{cases} \bar{x} & \text{auf } \mathcal{B}, \\ \frac{1}{n-1} \sum_{v \neq M} x_v = \bar{x} - \frac{1}{n-1} z_M & \text{auf } \bar{\mathcal{B}}. \end{cases}$$

Mit anderen Worten: Die am weitesten außen liegende Beobachtung wird dann nicht mit zur Schätzung des Erwartungswertes benutzt, wenn die absolute Differenz zum Stichprobenmittel größer ist als eine vorher festgelegte Konstante B . Dabei wird B meistens durch den Fehler erster Art α festgelegt; B ist eine monoton fallende Funktion von α .

Um zu untersuchen, ob sich das Risiko r'' von $\bar{x} - b(z)$ wesentlich vom Risiko r' der Bayeslösung $\bar{x} - a(z)$ unterscheidet, wurden beide Risiken für $n = 3, 4, 5$ und $\sigma = 3, 5, 7$ sowie für verschiedene Vorbewertungen φ bzw. Schranken B numerisch berechnet.

Die Integrale (6.4) wurden auf einer elektronischen Rechenmaschine IBM 650 ausgewertet, die freundlicherweise die Allianz Versicherungs - AG, München, zur Verfügung stellte. Das Ergebnis ist in Tabelle 1 zusammengestellt. Die Genauigkeit ist nicht groß; der Fehler beträgt bei r'_i ($i \neq 0$) etwa 2 bis 6%, bei r'_0 ein bis zwei Einheiten in der letzten Dezimale. Eine größere Genauigkeit hätte eine entsprechend längere Rechenzeit erfordert; in diesem Zusammenhang ist aber nur die ungefähre Größe des Risikos wichtig. Für $\varphi = 0$ ist $a(z) \equiv 0$, und das Integral läßt sich geschlossen auswerten.

Die Minimallösung ist in allen untersuchten Fällen die Bayeslösung zu $1 - n\varphi = 0$, $\varphi = 1/n$.

	n	σ	φ						
			$1/n$	0,15	0,10	0,05	0,025	0,01	0,00
r'_0	3	3	,031	,015	,011	,006	,004	,002	,000
		5	,044	,017	,012	,006	,003	,001	,000
		7	,050	,019	,013	,008	,005	,002	,000
	4	3	,021	,013	,009	,005	,003	,002	,000
		5	,030	,016	,011	,006	,004	,002	,000
		7	,029	,012	,007	,003	,002	,001	,000
	5	3	,015	,011	,007	,004	,002	,001	,000
		5	,020	,013	,008	,004	,003	,002	,000
		7	,021	,013	,007	,003	,002	,001	,000
r'_i $i \neq 0$	3	3	,23	,23	,24	,26	,28	,31	,748
		5	,23	,24	,25	,27	,30	,34	2,510
		7	,19	,20	,21	,23	,26	,30	5,172
	4	3	,095	,098	,102	,113	,126	,148	,429
		5	,096	,100	,105	,118	,133	,156	1,421
		7	,071	,075	,079	,090	,104	,125	2,919
	5	3	,064	,065	,067	,073	,081	,094	,277
		5	,055	,056	,059	,066	,076	,091	,912
		7	,051	,052	,054	,061	,068	,080	1,871

Tabelle 1.

Risiko r'_i der Bayeslösungen (6.3)

Um die Risiken r' und r'' sinnvoll vergleichen zu können, müssen wir natürlich für die Schätzfunktion $\bar{x} - b(z)$ den gleichen Schaden wie für $\bar{x} - a(z)$ benutzen, also analog zu (6.1)

$$d_0(z; b(z)) = b(z)^2,$$

$$d_i(z; b(z)) = \left[b(z) - \frac{1}{n \cdot K} z_i \right]^2 \quad (i \neq 0).$$

Dann gilt

$$\begin{aligned}
 (6.7) \quad r_0'' &= (2\pi)^{-n/2} \cdot n \int_{\mathcal{Z}} \left[\frac{1}{n-1} z_M \right]^2 \exp \left\{ -\frac{n}{2} (\bar{x} - m)^2 \right. \\
 &\quad \left. - \frac{1}{2} \sum z_v^2 \right\} d\bar{x} dz \\
 &= (2\pi)^{-(n-1)/2} \sqrt{n} \int_{|z_M| > B} \left[\frac{z_M}{n-1} \right]^2 \exp \left\{ -\frac{1}{2} \sum z_v^2 \right\} dz, \\
 r_1'' &= (2\pi)^{-n/2} \frac{n}{\sigma} \int_{\mathcal{Z}} \left[\frac{z_M}{n-1} - \frac{z_1}{nK} \right]^2 \exp \left\{ -\frac{k\sigma^{-2}}{2} (\bar{x}-m)^2 \right. \\
 &\quad \left. - \frac{1}{2} \sum z_v^2 + (1-\sigma^{-2})(\bar{x}-m)z_1 + \frac{1}{2}(1-\sigma^{-2})z_1^2 \right\} d\bar{x} dz \\
 &\quad + (2\pi)^{-n/2} \frac{n}{\sigma} \int_{\mathcal{Z}} \left[\frac{z_1}{nK} \right]^2 \exp \left\{ -\frac{k\sigma^{-2}}{2} (\bar{x}-m)^2 - \frac{1}{2} \sum z_v^2 \right. \\
 &\quad \left. + (1-\sigma^{-2})(\bar{x}-m)z_1 + \frac{1}{2}(1-\sigma^{-2})z_1^2 \right\} d\bar{x} dz \\
 &\quad (i \neq 0).
 \end{aligned}$$

Das über den R^n erstreckte Integral

$$(2\pi)^{-n/2} \frac{n}{\sigma} \int \frac{1}{n^2 K^2} z_1^2 f_1(\bar{x}, z) d\bar{x} dz \quad (i \neq 0)$$

kann geschlossen ausgewertet werden und ergibt

$$\frac{(n-1)(\sigma^2-1)}{n^3 K}. \text{ Die Integration über } \bar{x} \text{ liefert somit}$$

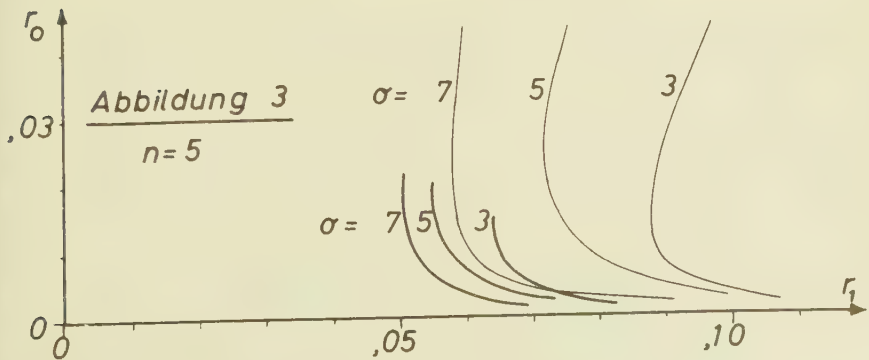
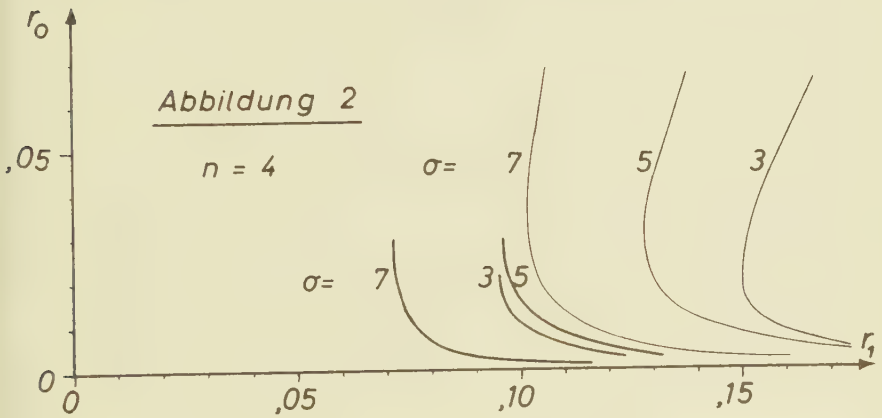
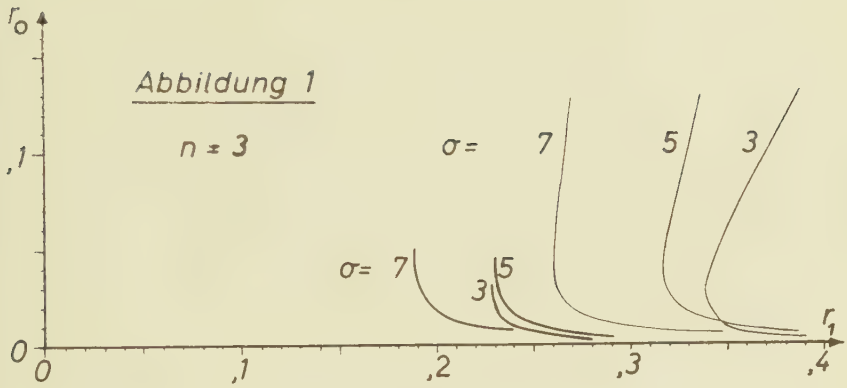
$$\begin{aligned}
 (6.7 \text{ a}) \quad r_1'' &= \frac{(n-1)(\sigma^2-1)}{n^3 K} + \frac{n}{\sqrt{k}} (2\pi)^{-(n-1)/2} \int_{|z_M| > B} \left[\frac{z_M^2}{(n-1)^2} \right. \\
 &\quad \left. - 2 z_1 \frac{z_M}{nK(n-1)} \right] \exp \left\{ -\frac{1}{2} \sum z_v^2 + \frac{z_1^2}{2K} \right\} dz \\
 &\quad (i \neq 0).
 \end{aligned}$$

In Tabelle 2 sind die Ergebnisse der Integrale (6.7) bzw. (6.7 a) für verschiedene Werte von n , σ und B angegeben, dazu auch für $B \geq 2$ der ungefähre Fehler erster Art α . Der Fehler der numerischen Rechnungen dürfte bei r''_i ($i \neq 0$) für $n=3$ sowie für $n=4$, $B \geq 2,0$ und $n=5$, $B \geq 2,5$ höchstens 0,002 betragen, desgleichen bei r''_0 für alle Werte von n und B . Für $n=4$, $B=1,5$ ist der Fehler auf alle Fälle kleiner als 0,008; für $n=5$, $B \leq 2,0$ ist eine brauchbare Abschätzung nur sehr mühsam zu erhalten.

n	B	α	r''_0	r''_i		
				$\sigma = 3$	$\sigma = 5$	$\sigma = 7$
3	1,5		,135	,394	,338	,271
	2,0	,038	,049	,346	,319	,261
	2,5	,0062	,012	,344	,345	,287
	3,0	,0008	,002	,391	,432	,373
	∞	,0000	,000	,748	2,510	5,172
4	1,5		,106	,187	,145	,111
	2,0	,076	,045	,158	,130	,102
	2,5	,015	,013	,153	,137	,110
	3,0	,0021	,003	,179	,176	,146
	∞	,0000	,000	,429	1,421	2,919
5	1,5		,082	,115	,086	,064
	2,0	,12	,039	,095	,074	,059
	2,5	,025	,012	,088	,076	,059
	3,0	,0039	,003	,102	,096	,078
	∞	,0000	,000	,277	,912	1,871

Tabelle 2.

Risiko r''_0 und r''_i ($i \neq 0$) der Schätzfunktion (6.6)



Wir wollen nun die Risiken r' und r'' der Schätzfunktionen $\bar{x} - a(z)$ bzw. $\bar{x} - b(z)$ miteinander vergleichen. Dazu haben wir in den Abbildungen 1 bis 3 r_0 gegen r_1 aufgetragen. Die stark ausgezogenen Linien stellen das Risiko r' der Bayeslösungen dar und bilden zugleich einen Teil des Randes des Bereiches $\mathcal{R}$ aller Risikovektoren. Die dünnen Linien stammen von den Risiken r'' der Schätzfunktion $\bar{x} - b(z)$ und liegen im Innern von $\mathcal{R}$.

Wie man sieht, sind r' und r'' von der gleichen Größenordnung. Wählt man beispielsweise $r'_0 = r''_0$, so ist r'_1 nur etwa $1 \frac{1}{3}$ bis $1 \frac{1}{2}$ mal so groß wie r''_1 . Je größer n und σ sind, desto geringer werden die Unterschiede zwischen r' und r'' . Bei $n = 5$, $\sigma = 7$ liegen beide Kurven bereits recht dicht beieinander.

Diese Tatsache läßt es gerechtfertigt erscheinen, in der Praxis die Schätzfunktion $\bar{x} - b(z)$ der wesentlich komplizierter gebauten Bayeslösung $\bar{x} - a(z)$ vorzuziehen. Außerdem hat $\bar{x} - b(z)$ noch den Vorteil, nicht vom Ausreißerparameter σ abzuhängen, der bei $a(z)$ in den Konstanten C und K enthalten ist. - Das Risiko hängt natürlich bei beiden Schätzfunktionen von σ ab.

Zum Abschluß sei noch ohne Beweis die Schätzfunktion $\bar{x} - a_1(z)$ angeführt, die man in der gleichen Weise wie $\bar{x} - a(z)$ erhält, wenn mehr als ein Ausreißer zugelassen ist und jede Hypothese, die r Ausreißer annimmt, die Vorbewertung ψ_r erhält. Wir benutzen die Bezeichnungen

$$S := 1 - \sigma^{-2},$$

$$h(i_1, \dots, i_r) := \exp \left\{ \frac{S}{2} \sum_{i=1}^r z_i^2 + \frac{1}{2} \frac{S^2}{n - rS} \left[\sum_{i=1}^r z_i \right]^2 \right\}.$$

Außerdem soll $(i_1, \dots, i_r)$ als Summationsbuchstabe

eine Summation über alle r -Tupel $i_1 < i_2 < \dots < i_r$ bedeuten. Dann ergibt sich

$$a_1(z) = \frac{\sum_{r=1}^{n-1} \frac{\psi_r \cdot \sigma^{-r}}{(n-rS)^{3/2}} \sum_{(i_1, \dots, i_r)} \left[h(i_1, \dots, i_r) \sum_{i=1}^{i_r} z_i \right]}{\sum_{r=0}^n \frac{\psi_r \cdot \sigma^{-r}}{(n-rS)^{1/2}} \sum_{(i_1, \dots, i_r)} h(i_1, \dots, i_r)} .$$

(6.3) ist der Spezialfall dieser Gleichung für

$$\psi_0 = 1 - n\varphi, \quad \psi_1 = \varphi, \quad \psi_2 = \dots = \psi_n = 0.$$

Literaturverzeichnis

Abkürzungen:

AMSt Annals of Mathematical Statistics
Bka Biometrika

a) Literatur zum Ausreißerproblem

- [1] C.I.Bliss, W.G.Cochran und J.W.Tukey, A rejection criterion based upon the range, Bka 43 (1956) 418 - 422
- [2] W.G.Cochran, The distribution of the largest of a set of estimated normal variances as a fraction of their total, Ann. of Eugenics 11 (1941) 47 - 52
- [3] H.A.David, On the application to statistics of an elementary theorem in probability, Bka 43 (1956) 85 - 91
- [4] W.J.Dixon, Analysis of extreme values, AMSt 21 (1950) 488 - 506
- [5] W.J.Dixon, Ratios involving extreme values, AMSt 22 (1951) 68 - 78
- [6] R.Doornbos, Significance of the smallest of a set of estimated normal variances, Statistica Nederlandica 10 (1956) 17 - 26
- [7] R.Doornbos und H.J.Prins, Slippage tests for a set of gamma variates, Indagationes Mathematicae 18 (1956) 329 - 337

- [8] R.Doornbos und H.J.Prins, On slippage tests, Proc. Kon. Ned. Akad. van Wetensch. 61 Series A (1958) 38 - 55, 438 - 447
- [9] H.J.Godwin, On the distribution of the estimate of mean deviation obtained from samples from a normal population, Bka 33 (1945) 254-256
- [10] F.E.Grubbs, Sample criteria for testing outlying observations, AMSt 21 (1950) 27 - 58
- [11] M.Halperin, S.W.Greenhouse, J.Cornfield und J.Zalokar, Tables of percentage points for the studentized maximum absolute deviate in normal samples, Journal of the Amer. Statist. Ass. 50 (1955) 185 - 195
- [12] H.O.Hartley, The range in random samples, Bka 32 (1942) 334 - 348
- [13] H.O.Hartley, Note on the calculation of the distribution of the estimate of mean deviation in normal samples, Bka 33 (1945) 257 - 258
- [14] J.O.Irwin, On a criterion for the rejection of outlying observations, Bka 17 (1925) 238 - 250
- [15] M.N.Kapur, A property of the optimum solution suggested by Paulson for the k - sample slippage problem for the normal distribution, J. Indian Soc. Agric. Statist. 2 (1957) 179 - 190
- [16] W.H.Kruskal, A nonparametric test for the several sample problem, AMSt 23 (1952) 525 - 540
- [17] A.Kudo, On the testing of outlying observations, Sankhyā 17 (1956/57) 67 - 76
- [18] J.M.May, Extended and corrected tables of the upper percentage points of the studentized range, Bka 39 (1952) 192 - 193

- [19] A.T.McKay, The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe, Bka 27 (1935) 466 - 471
- [20] F.Mosteller, A k -sample slippage test for an extreme population, AMSt 19 (1948) 58 - 65
- [21] F.Mosteller und J.W.Tukey, Significance levels for a k -sample slippage test, AMSt 21 (1950) 120 - 123
- [22] K.R.Nair, The distribution of the extreme deviate from the sample mean and its studentized form, Bka 35 (1948) 118 - 144
- [23] D.Newman, The distribution of ranges in samples from a normal population, expressed in terms of an independent estimate of the standard deviation, Bka 31 (1940) 20 - 30
- [24] J.Pachares, Table of the upper 10% points of the studentized range, Bka 46 (1959) 461 - 466
- [25] E.Paulson, An optimum solution to the k -sample slippage problem for the normal distribution, AMSt 23 (1952) 610 - 616
- [26] E.S.Pearson, A further note on the distribution of range in samples taken from a normal population, Bka 18 (1926) 173 - 194
- [27] E.S.Pearson, The percentage limits for the distribution of range in samples from a normal population, Bka 24 (1932) 404 - 417
- [28] E.S.Pearson und C.ChandraSekar, The efficiency of statistical tools and a criterion for the rejection of outlying observations, Bka 28 (1936) 308 - 320

- [29] E.S.Pearson und H.O.Hartley, The probability integral of the range in samples of n observations from the normal population, Bka 32 (1942) 301 - 310
- [30] E.S.Pearson und H.O.Hartley, Tables of the probability integral of the studentized range, Bka 33 (1943) 89 - 99
- [31] E.S.Pearson und H.O.Hartley, Tables of the probability integral of the mean deviation in normal samples, Bka 33 (1945) 259 - 265
- [32] J.Pfanzagl, Ein kombiniertes Test- und Klassifikationsproblem, Metrika 2 (1959) 11 - 45
- [33] K.C.S.Pillai, On the distribution of the studentized range, Bka 39 (1952) 194 - 195
- [34] K.C.S.Pillai, Upper percentage points of the extreme studentized deviate from the sample mean, Bka 46 (1959) 473 - 474
- [35] K.V.Ramachandran und C.G.Khatri, On a decision procedure based on the tukey statistic, AMSt 28 (1957) 802 - 806
- [36] S.Rosenbaum, Tables for a nonparametric test of dispersion, AMSt 24 (1953) 663 - 668
- [37] N.V.Smirnoff, On the estimation of the maximum term in a series of observations, Doklady Akad. Nauk SSSR, New Series 33 (1941) 346 - 350
- [38] "Student", Errors of routine analysis, Bka 19 (1927) 151 - 164
- [39] W.R.Thompson, On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation, AMSt 6 (1935) 214 - 219

- [40] L.H.C.Tippett, On the extreme individuals and the range of samples taken from a normal population, Bka 17 (1925) 364 - 387
- [41] D.R.Truax, An optimum slippage test for the variance of k normal distributions, AMSt 24 (1953) 669 - 674
- [42] J.E.Walsh, Large sample nonparametric rejection of outlying observations, Ann. Inst. Statist. Math. Tokyo 10 (1959) 223 - 232

Tafelwerke:

- [43] Biometrika Tables for Statisticians, edited by E.S.Pearson and H.O.Hartley, Vol. I, Cambridge 1954
- [44] Tables for Statisticians and Biometricians, Vol. II, edited by K.Pearson, Cambridge 1931

b) Literatur zu § 3

- [45] D.Blackwell, The range of certain vector integrals, Proc. Amer. Math. Soc. 2 (1951) 390 - 395
- [46] E.L.Lehmann, Some principles of the theory of testing hypotheses, AMSt 21 (1950) 1 - 26
- [47] E.L.Lehmann, Testing Statistical Hypotheses, New York und London 1959
- [48] D.V.Lindley, Statistical inference, Journal of the Royal Statist. Soc., Series B, 15 (1953) 30 - 76
- [49] H.Richter, Wahrscheinlichkeitstheorie, Berlin, Göttingen und Heidelberg 1956

- [50] H.Richter, Über optimale mehrstufige Tests, Transactions of the 2nd Prague Conference on Information Theory, Statist. Decision Functions, Random Processes (1960) 557-568
- [51] A.Wald, Statistical Decision Functions, New York und London 1950

L e b e n s l a u f

Ich wurde am 26. Juni 1933 in Ober - Bellmannsdorf, Kreis Lauban, als Sohn des Pfarrers Dr. Friedrich Gebhardt und seiner Ehefrau Eva, geb. Steinhoff, geboren. Im März 1952 legte ich an der Städtischen Oberschule Wolfsburg das Abitur ab. Ich studierte zwei Semester an der Technischen Hochschule Braunschweig und je vier Semester an den Universitäten Hamburg und München. Im Februar 1957 legte ich an der Universität München bei Herrn Prof. Dr. W.Maak die Diplomprüfung für Mathematik ab. Von März 1957 bis September 1960 arbeitete ich im Rechenzentrum der Allianz Versicherungs - AG, München; längere Zeit davon war ich zur Fertigstellung der Dissertation beurlaubt. Seit November 1960 bin ich in der Statistischen Arbeitsgruppe von Herrn Prof. Dr. H.Kneser, Tübingen, tätig.

Ich hörte u.a. Vorlesungen bei Frau Professor H.Braun und den Herren Professoren und Dozenten G.Aumann, F.Bopp, L.Collatz, A.Güttler, H.Hasse, O.Heckmann, W.Heisenberg, W.Klingenberg, M.Kneser, W.Lenz, W.Maak, H.Richter, H.Röhr1, B.Schoeneberg, E.Sperner, K.Stein und A.A.Wachmann.

